World Scientific Series in 20th Century Physics – Vol. 5

Selected Papers of ABDUS SALAM

(WITH COMMENTARY)

Editors

A. Ali
C. Isham
T. Kibble
Riazuddin

World Scientific

Selected Papers of ABDUS SALAM

(WITH COMMENTARY)

World Scientific Series in 20th Century Physics

Published

Vol. 1 Gauge Theories – Path and Future
edited by R. Akhoury, B. de Wit and P. van Nieuwenhuizen

Vol. 2 Scientific Highlights in Memory of Léon van Hove
edited by F. Nicodemi

Forthcoming

Vol. 3 Selected Papers, with Commentary, of T. H. R. Skyrme
edited by G. E. Brown

Vol. 4 Salamfestschrift
edited by A. Ali, J. Ellis and S. Randjbar-Daemi

Vol. 6 Research on Particle Detectors
by G. Charpak

Vol. 7 A Career in Theoretical Physics
by P. W. Anderson

World Scientific Series in 20th Century Physics – Vol. 5

Selected Papers of ABDUS SALAM

(WITH COMMENTARY)

Editors

A. Ali
DESY

C. Isham
Imperial College of Science and Technology

T. Kibble
Imperial College of Science and Technology

Riazuddin
King Fahd University of Petroleum and Minerals

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 73 Lynton Mead, Totteridge, London N20 8DH

Library of Congress Cataloging-in-Publication Data

Salam, Abdus, 1926–

Selected papers of Abdus Salam : with commentary / editors, A. Ali . . . [et al.].

p. cm. -- (World scientific series in 20th century physics ; vol. 5)

Includes bibliographical references.

ISBN 9810216629 9810216637 (pbk)

1. Particles (Nuclear physics). 2. Electroweak interactions. 3. Grand unified theories (Nuclear physics). 4. Quantum field theory. I. Ali, A. (Ahmed) II. Title. III. Series.

QC793.2.S25 1994

530.1'43--dc20 94–2323

CIP

The editors and publisher would like to thank the following publishers of the various journals and books for their assistance and permissions to include the selected reprints found in this volume:

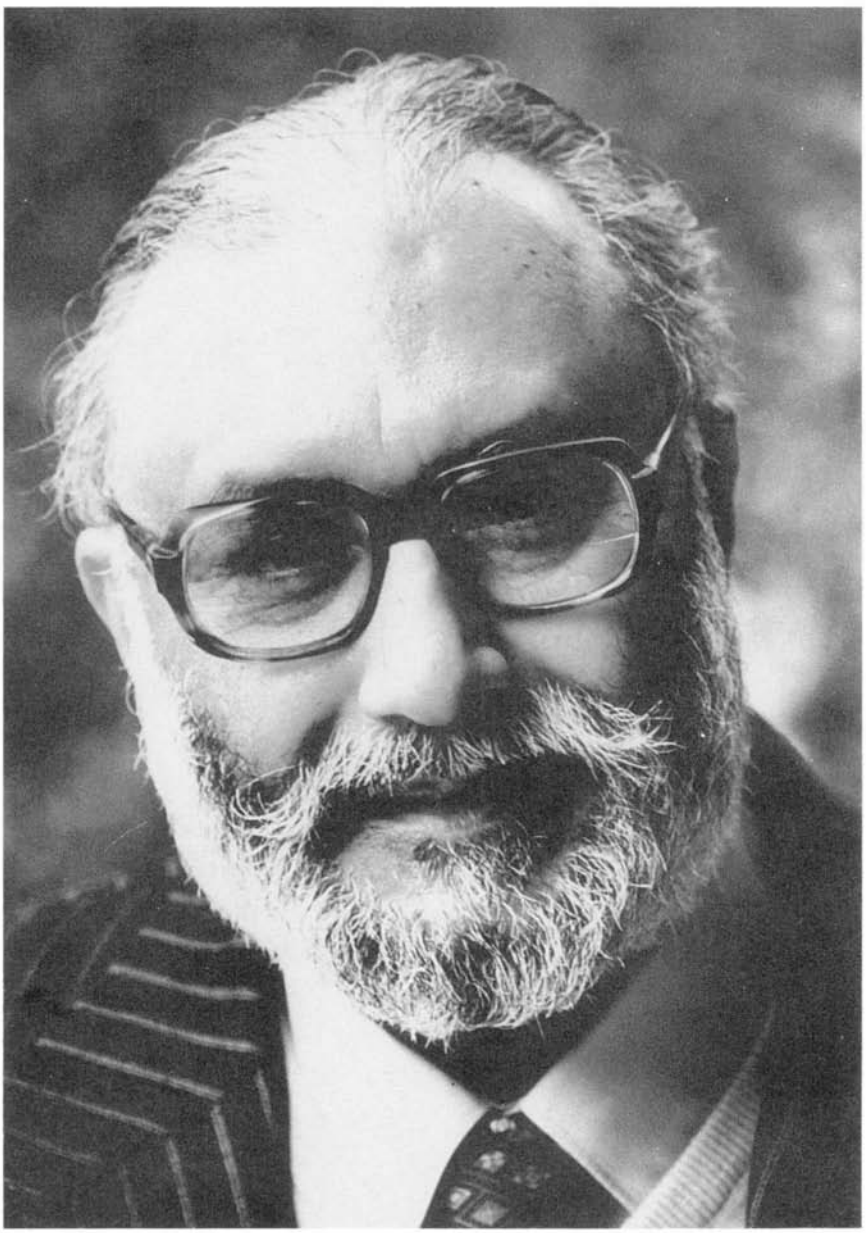
Academic Press
American Physical Society
Cambridge University Press
Elsevier Science Publishers B.V.
John Wiley and Sons
Macmillan Magazines Ltd.
The Nobel Foundation
Royal Soc. London
Società Italiana di Fisica
Springer Verlag

Copyright © 1994 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA.

Printed in Singapore by Singapore National Printers Ltd



This page is intentionally left blank

Introduction

Abdus Salam is one of the leading theoretical physicists of his generation. His 275 published papers span a huge range and contain many seminal contributions. It is quite impossible to include all the important papers in a single volume. As editors, we have had to make very difficult choices, omitting many papers that we would have wished to accommodate. We have tried to include representatives of all the different periods of his work and all the different fields to which he contributed, though inevitably some are better represented than others.

The papers appear in roughly chronological order, but they have been grouped into five broad sections, each representing a different major strand in his work.

Abdus Salam was born in Jhang, now in Pakistan, one of the least developed areas of a developing country. Before leaving for England, his name was already a legend in his home country as, in the fierce competition of pre-partition India, he passed every examination at Punjab University, Lahore with the creation of a new record.

Salam came to Cambridge in 1946. Following a brilliant undergraduate career, he began research in the Cavendish Laboratory in 1949 at a most opportune moment. During the second world war, many of the leading theoretical physicists had been involved in the Manhattan project or other work for the war effort. When the war ended they returned to their universities to attack with renewed vigour the problems they had left behind, including the intractable problem of infinities in quantum electrodynamics (QED). There was a sudden surge of progress, most notably the development of renormalization theory, which showed how to control the infinities by absorbing them into corrections to the observable parameters of the theory, the mass and charge of the electron.

When Salam graduated, he asked Nicholas Kemmer to take him on as a research student. Kemmer told him that all the fundamental problems of making QED finite had been solved by Shin-ichiro Tomonaga, Julian Schwinger, Richard Feynman and Freeman Dyson, and that Paul Matthews, then just completing his PhD, had nearly done the same for spin-zero meson theory, but might have some problems left. Matthews had proved renormalizability to lowest order and suggested that Salam try to extend the proof to higher orders. He was amazed when Salam returned almost at once with a solution to one of the outstanding difficulties, the problem of 'overlapping divergences'. Matthews became Salam's PhD supervisor and they worked together on the problem. Thus began one of the longest-lasting and most productive collaborations of Salam's career.

Salam's work made an immediate impact. He was at once recognized as a major contributor to the field. Following a brief visit to Princeton and a three-year spell as professor at his *alma mater*, Government College, Lahore, and Head of the Mathematics Department in Panjab University, he returned to Cambridge as a lecturer in 1954 and was appointed Professor of Theoretical Physics at Imperial College in 1957 at the age of 30. During this period he made some of the key advances in quantum field theory, especially concerning proofs of renormalizability and extending the validity of dispersion relations. These papers constitute the bulk of the first section of the collection.

Following the successful description of electromagnetic interactions by QED, physicists sought to apply similar ideas to the other fundamental interactions — the strong and weak nuclear forces and gravity. Salam made one of the key early advances in developing a theory of weak interactions when he introduced the γ_5 invariance principle for neutrinos, leading to the two-component theory of the neutrino. Since the triumph of QED, the greatest advance in fundamental theoretical physics has been the development in the 60s and 70s of the unified theory of weak and electromagnetic interactions, a gauge theory based on the symmetry group SU(2)×U(1). Abdus Salam played a major role, recognized by his sharing the 1979 Nobel Prize for Physics with Sheldon Glashow and Steven Weinberg. Salam's contributions in this field are represented by the papers in Sec. 2.

Once again this success was a spur to further work, a challenge to extend still further the ideas of unification. The next two sections of the book represent two different directions in which Salam sought to go beyond the electroweak theory.

One obvious question was: is it possible to develop a unified theory that embraces not only electromagnetic and weak interactions, but the strong interactions too? By this time, physicists had a good understanding of both electroweak and strong interactions separately. Both are described by gauge theories — in the case of the strong interactions, by quantum chromodynamics (QCD), based on the symmetry group SU(3). According to this theory the strongly interacting particles, or hadrons, are seen as composed of quarks, held together by exchange of gluons. Thus the proton and neutron are each made of three quarks, and the pion of a quark-antiquark pair. So theorists started to look for a 'grand unified theory' that would combine the symmetries of the strong and electroweak theories, SU(3)×SU(2)×U(1), into a single larger group, a symmetry that would unite the quarks and the leptons (the electron, muon, tauon and neutrinos). One such proposal, due to Jogesh Pati and Salam, is described in the papers collected in Sec. 3. This was an influential and fascinating development.

The greatest remaining challenge to theoretical physics is to bring gravity within the same framework as the other interactions. So far, a viable quantum theory of

gravity eludes us. Though we have an excellent classical theory of gravity — Albert Einstein's general theory of relativity — it cannot be quantized; the attempt leads to irremovable infinities.

One possible direction of escape is to change the classical theory, replacing general relativity by 'supergravity' or 'superstring theory'. One of the key ideas here is *supersymmetry*, which relates fermions to bosons. Though there is no direct experimental evidence for supersymmetry, there are indirect hints, and it is theoretically a most attractive hypothesis, which does much to tame the infinities. Even more radically, string theory postulates that the most fundamental entities are not point-like but extended, one-dimensional objects. This too helps to tame the infinities. It is possible that by some combination of these ideas we will eventually reach a viable theory, not only embracing a quantum theory of gravity, but also uniting all four interactions in a 'theory of everything' — though this still remains a distant hope. Section 4 contains a number of major contributions by Salam and his collaborators to these developments.

The final section, devoted to some of Salam's most recent work, illustrates both the breadth of his interests and the continuing originality of his ideas. Unlike most of his previous research, these papers are not concerned with elementary particles but with fundamental problems in condensed matter physics and biology.

One of the most exciting developments in the condensed matter field in recent years was the discovery of materials that are superconducting up to relatively high temperatures. Whereas the mechanism of superconductivity in conventional metallic superconductors is now well understood, there is as yet no accepted theory for the mechanism of high-temperature superconductivity. The papers presented here describe a very promising line of attack, involving field theories restricted to two spatial dimensions, corresponding to the layered structure of these materials.

Also included are two papers proposing a solution to one of the long-standing puzzles of evolutionary biology — the origin of chirality or handedness in biological molecules.

Each of the five sections is preceded by a brief introduction, setting the scene for the papers that follow.

The editors are very conscious of the omissions from this volume. Many important papers, especially long ones, had to be eliminated to keep the total number of pages within bounds. However, a complete list of Salam's published papers is included from which those we could not accommodate can be found.

The volume also omits all Salam's many wider contributions — to debates on science policy, development in the third world, international peace, and so on. Many of these have already appeared in other collections.

We are grateful to numerous colleagues for their advice about what should be included, and only regret that we were unable to accept all their suggestions.

Ahmed Ali
Christopher Isham
Tom Kibble
Riazuddin

Contents

Introduction vii

1. Quantum Field Theory and Dispersion Relations

[1]* A Problem of Ramanujam 6

[3] Overlapping Divergences and the S-Matrix 8

[4] The Renormalization of Meson Theories (with P. T. Matthews) 19

[5] Divergent Integrals in Renormalizable Field Theories 23

[8] Renormalized S-Matrix for Scalar Electrodynamics 29

[12] Fredholm Theory of Scattering in a Given Time-Dependent Field (with P. T. Matthews) 43

[15] Covariant Fock Equations (with P. T. Matthews) 49

[16a] Renormalization (with P. T. Matthews) 56

[17] The Bound State Representation (with J. Hamilton) 63

[18] On the Relativistic Equation for Scattering (with N. Kemmer) 73

[22] On Generalized Dispersion Relations. I 79

[24] On Generalized Dispersion Relations. II (with W. Gilbert) 85

[37] Relativistic Field Theory of Unstable Particles (with P. T. Matthews) .. 90

[42] Relativistic Theory of Unstable Particles. II (with P. T. Matthews) 95

[50] Nonleptonic Decay Modes of the Hyperons (with G. Feldman and P. T. Matthews) 101

[53] Change of Variables and Equivalence Theorems in Quantum Field Theories (with S. Kamefuchi and L. O'Raifeartaigh) 103

[65] Renormalizable Electrodynamics of Vector Mesons 124

[113] Momentum-Space Behavior of Integrals in Nonpolynomial Lagrangian Theories (with J. Strathdee) 128

*Numbers in brackets refer to the List of Publications in Chronological Order on pp. 663–679.

[144] Comment on the Computation of Effective Potentials (with J. Strathdee) 145

2. Symmetries and Electroweak Unification

[21] On the Classification of Fundamental Particles (with J. C. Polkinghorne) 161

[28] On Parity Conservation and Neutrino Mass 167

[30] On Fermi Interactions 170

[32] On Symmetries of Weak Interactions (with B. D'Espagnat and J. Prentki) 175

[39] Weak and Electromagnetic Interactions (with J. C. Ward) 183

[48] $\Delta I = 1/2$ Rule (with J. C. Ward) 193

[49] On a Gauge Theory of Elementary Interactions (with J. C. Ward) 195

[58] Isotopic Spin Relations in Hyperon Production (with J. Tiomno) 201

[64] Broken Symmetries (with J. Goldstone and S. Weinberg) 204

[72] Electromagnetic and Weak Interactions (with J. C. Ward) 210

[78] Relativistic Structure of SU(6) (with R. Delbourgo and J. Strathdee) 214

[80] The Covariant Theory of Strong Interaction Symmetries (with R. Delbourgo and J. Strathdee) 218

[94] Harmonic Analysis in Terms of the Homogeneous Lorentz Group (with R. Delbourgo and J. Strathdee) 231

[99] Reggeization of Quark Number (with R. Delbourgo and J. Strathdee) . 234

[101] Weak and Electromagnetic Interactions 244

[104] Finite-Component Field Representations of the Conformal Group (with G. Mack) 255

[106] Nonlinear Realizations. I. The Role of Goldstone Bosons (with J. Strathdee) 284

[169] Transition to CP Conservation and Zero Cabibbo Angle in Strong Magnetic Fields (with J. Strathdee) 294

[176] Intensity-Dependent Mass Shift and Symmetry Breaking (with T. W. B. Kibble and J. Strathdee) 297

[216] Gauge Unification of Fundamental Forces 306

3. Lepton-Hadron Unification

[152] Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions (with J. C. Pati) 327

[150] Is Baryon Number Conserved? (with J. C. Pati) 339

[155] Lepton Number as the Fourth "Color" (with J. C. Pati) 343

[195a] Neutral-Current Phenomena within the Left-Right-Symmetric Unified Theory of Quarks and Leptons (with J. C. Pati and S. Rajpoot) 358

[256] The Moon as the Searching Ground for Proton Decay (with J. C. Pati and B. V. Sreekantan) 378

4. Gravity, Supersymmetry and Strings

[122] Infinity Suppression in Gravity-Modified Quantum Electrodynamics (with C. J. Isham and J. Strathdee) 389

[124] The Gravitational Correction to PCAC (with R. Delbourgo) 402

[156] Super-Gauge Transformations (with J. Strathdee) 404

[157] Unitary Representations of Super-Gauge Symmetries (with J. Strathdee) 410

[158] On Goldstone Fermions (with J. Strathdee) 417

[159] Supersymmetry and Non-Abelian Gauges (with J. Strathdee) 420

[160] Superfields and Fermi-Bose Symmetry (with J. Strathdee) 423

[165] Feynman Rules for Superfields (with J. Strathdee) 438

[187] Black Holes as Solitons (with J. Strathdee) 449

[228] On Kaluza-Klein Theory (with J. Strathdee) 451

[235] Spontaneous Compactification in Six-Dimensional Einstein-Maxwell Theory (with S. Randjbar-Daemi and J. Strathdee) 488

[241] Instanton Induced Compactification and Fermion Chirality (with S. Randjbar-Daemi and J. Strathdee) 510

[251] An Anomaly-Free Model in Six Dimensions (with S. Randjbar-Daemi, E. Sezgin and J. Strathdee) 515

[254] Locally Supersymmetric σ-Model with Wess-Zumino Term in Two Dimensions and Critical Dimensions for Strings (with E. Bergshoeff, S. Randjbar-Daemi, H. Sarmadi and E. Sezgin) 521

[263] σ-Models and String Theories (with S. Randjbar-Daemi and J. Strathdee) 541

[244] On Kaluza-Klein Cosmology (with S. Randjbar-Daemi and J. Strathdee) 568

5. Condensed Matter and Biology

[269] Chern-Simons Superconductivity at Finite Temperature (with S. Randjbar-Daemi and J. Strathdee) 575

[270] Anyons and Chern-Simons Theory on Compact Spaces of Finite Genus (with S. Randjbar-Daemi and J. Strathdee) 620

[271] The Role of Chirality in the Origin of Life 626

[274] Chirality, Phase Transitions and Their Induction in Amino Acids 635

[275] Generalized Spin Systems and σ-Models (with S. Randjbar-Daemi and J. Strathdee) 643

Biodata 659

List of Publications in Chronological Order 663

Selected Papers of ABDUS SALAM

(WITH COMMENTARY)

This page is intentionally left blank

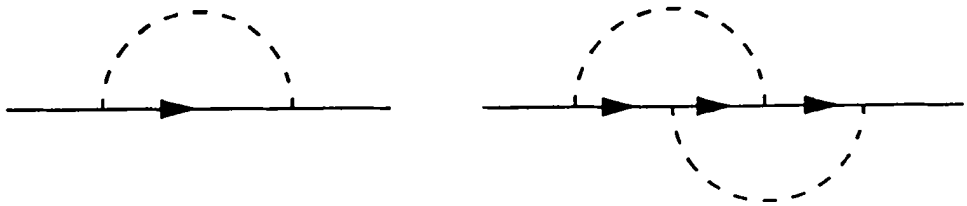
1. Quantum Field Theory and Dispersion Relations

The first paper in this collection is a remarkable contribution from a 17-year-old student at Government College, Lahore, in which he describes a better way of solving a set of coupled nonlinear algebraic equations than the one found earlier by the mathematical genius Srivanasan Ramanujan. It shows that from a very early age Abdus Salam exhibited astonishing mathematical ingenuity and originality. It is of course a mere foretaste of what was to come. Salam's other early contributions, of which only a small selection is reprinted here, are far more significant. He played a major role in the epoch-making advances in quantum field theory from 1950 onwards.

Soon after the advent of quantum mechanics, Paul Dirac and others applied it to the electromagnetic field, but they soon ran into difficulties. The first-order perturbation theory results gave good agreement with experiment, but all attempts to go beyond the leading order yielded nonsensical infinite answers. It was only after the second world war that Julian Schwinger, Shin-ichiro Tomonaga, Richard Feynman, and Freeman Dyson showed how the infinities could all be absorbed into corrections to the mass and charge of the electron. This 'renormalization theory' at last allowed physicists to get sensible answers for higher-order corrections. Within a very few years, quantum electrodynamics (QED) became one of the most accurately tested of all scientific theories, with successful predictions of radiative corrections to such quantities as the magnetic moment of the electron and the Lamb shift in the energy levels of hydrogen.

The problem Paul Matthews had worked on was whether 'meson theories' involving interactions between scalar, pseudoscalar or vector meson fields and spinor fields, were also renormalizable. He had shown that scalar and pseudoscalar theories with Yukawa interactions, represented by an interaction Lagrangian of the form $g\bar{\psi}\psi\phi$ or $g\bar{\psi}i\gamma_5\psi\phi$, could be renormalized at least in lowest order in the coupling constant g. He suggested that Salam try to extend the proof to higher orders. Almost immediately, Salam succeeded in solving one of the outstanding difficulties, that of 'overlapping divergences'. The problem arises with diagrams like that of Fig. 1b, which contains two overlapping subdiagrams, each of the form of Fig. 1a. Both the corresponding integrals must be rendered finite, but it is not obvious how to do this simultaneously. Salam, as a very young research student in Cambridge, deployed his formidable mathematical skills to bridge this gap, thus completing Dyson's proof that QED could be rendered entirely finite by renormalization.

From this auspicious start, Salam, together with his PhD supervisor, Paul Matthews, went on to prove that the renormalization procedure does indeed work



(a) (b)

Figure 1: Feynman diagrams representing (a) a primitive divergence and (b) overlapping divergences

for a variety of different theories involving interactions between scalar, pseudoscalar and spinor field and the electromagnetic field. Only a few of their many papers on this subject can be represented here.

Salam and Matthews also made a number of other very important contributions to the mathematical foundations of quantum field theory. In particular, they established connections between solutions obtained by Feynman's diagrammatic techniques and the methods used for the solution of Fredholm integral equations, thus bringing to bear much prior mathematical expertise. They also derived interesting and useful results concerning the equations satisfied by the Green's functions and 'Feynman amplitudes' — the vacuum expectation values of time-ordered products, such as (in a scalar theory),

$$\langle 0|T[\phi(x_1)\dots\phi(x_n)]|0\rangle\,. \tag{1}$$

In two papers which naturally followed their paper on the Fredholm solution, Salam and Matthews gave the first field-theoretic formulation of Feynman's 'sum-over-histories' principle.

A somewhat later paper worthy of special note is by Susumu Kamefuchi, Lochlainn O'Raifeartaigh and Salam. They addressed a question of considerable intrinsic interest and wide applicability, namely: when are two classically equivalent theories also equivalent quantum mechanically? Roy Chisholm had proved a theorem to the effect that if we take any quantum field theory and apply to it a 'point transformation' of the field variables — i.e., the new variables $\phi'(x)$ are local functions of the old $\phi(x)$ — then, provided the free Lagrangian is unchanged, the S-matrix elements are also unchanged. Kamefuchi, O'Raifeartaigh and Salam gave an improved version of the proof (for a slightly restated theorem), and discussed more generally the whole question of equivalence theorems relating different quantum field theories.

Following the triumph of QED, the outstanding unsolved problem was to find a viable theory of the strong interactions that bind protons and neutrons in atomic nuclei. The most popular idea was that strong interactions were carried by a triplet of fundamental pion fields, (π^+, π^0, π^-), represented as a vector in 'isospin space', $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$, where $\pi^{\pm} = (1/\sqrt{2})(\pi_1 \pm i\pi_2)$ and $\pi^0 = \pi_3$. The basic strong interaction was assumed to be a Yukawa interaction of the form $g\bar{\Psi}\boldsymbol{\tau}\Psi\cdot\boldsymbol{\pi}$, where Ψ is an eight-component nucleon field, combining the proton and neutron spinor fields,

$$\Psi = \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix}, \tag{2}$$

while $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$ denotes the Pauli matrices in isospin space.

The obvious difficulty with this theory was that in a strongly interacting theory, the coupling constant g, unlike the electric charge, would not be small. In QED, successive terms in the perturbation series are suppressed by increasing powers of the fine structure constant,

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137}, \tag{3}$$

but in a strong-interaction theory the corresponding factor would be of order 1 or larger, rendering perturbation theory useless.

Consequently, there was a widespread search for alternative methods of calculation, based so far as possible on fundamental properties of the theory such as relativistic invariance, causality and conservation of probability. A great deal of effort went into studying the implications of these fundamental requirements for the scattering amplitudes, or S-matrix elements, which encode the information obtainable from scattering experiments. One of the most successful approaches was based on dispersion relations, integral relations between the real and imaginary parts of scattering amplitudes, stemming from the causality requirement, analogous to the similar relations long used in optics and acoustics.

The requirement of causality can be shown to imply that the scattering amplitude $M(E)$ for an elastic scattering process is analytic in the complex energy plane, with a discontinuity across those parts of the real axis representing the energies of physically possible intermediate states. In the simplest case, it follows, by a contour-integral argument, that

$$\mathrm{Re}M(E) = \frac{\mathrm{P}}{\pi}\int_{-\infty}^{\infty} dE' \, \frac{\mathrm{Im}M(E')}{E' - E}, \tag{4}$$

where P denotes the principal part. In practice, the rate of decrease at infinity is usually not sufficient to ensure the validity of this simple relation; one needs to

include 'subtractions', for example to consider in place of M the subtracted function

$$\frac{M(E) - M(E_0)}{E - E_0} . \tag{5}$$

In 1955, Salam made a key contribution to the development of dispersion relations, extending the relations previously obtained for the specific case of forward scattering to arbitrary momentum transfer. Then with his student Walter Gilbert (later a distinguished molecular biologist), he enlarged their scope further to include spin-flip amplitudes. Later Matthews and Salam extended the discussion to cover the scattering of K-mesons, a case in which significant new features arise.

Dispersion relations were used, among other things, to establish connections between the strong and weak interactions. For example the Goldberger-Treiman relation between the weak axial vector constant and the strong pion-nucleon coupling, involving the leptonic decay rate of the pion, could be seen as arising from the pole term in a dispersion relation for a matrix element of the axial vector current. In an intriguing paper included in this collection, Gordon Feldman, Matthews and Salam discussed how far one could extend this to cover decays of the hyperons, Λ and Σ.

Another important fundamental contribution also stemming from the collaboration between Salam and Matthews was their treatment of unstable particles. They pointed out that the conventional approach was inconsistent with fundamental properties of relativistic quantum mechanics. An unstable particle is often represented as having a complex mass, $M - \frac{1}{2}i\lambda$, so that its wave function has a time dependence

$$\psi(\tau) \propto \exp(-iM\tau - \tfrac{1}{2}\lambda\tau) , \tag{6}$$

where τ is the proper time, and its probability changes exponentially, $|\psi|^2 \propto e^{-\lambda\tau}$. But although such a treatment may sometimes give the right answer, it cannot be fundamentally correct — energies cannot *really* be complex. In an S-matrix treatment, only truly stable particles can appear in the initial and final states; unstable particles appear only as resonance peaks in scattering amplitudes.

Salam and Matthews showed how one could set up a rigorous formalism in which the unstable particle states are treated as appropriate superpositions of the states of their decay products. They pointed out, among other results, that the exponential decay law is of limited validity — for very long times additional power-law terms (coming from thresholds in the scattering process) take over.

Salam continued to work not only on specific quantum field theory models, such as the unified electroweak theory described in the following section, but also on fundamental general aspects of quantum field theory. Included in this first section are two significant later contributions, both written jointly with John Strathdee.

One dates from 1970. Our present standard model of elementary particle interactions (gravity aside) is based entirely on what are now known to be renormalizable gauge theories. But the proofs of renormalizability, by Gerard 't Hooft, and by Benjamin Lee and J. Zinn-Justin, came later. In 1970 it still seemed highly likely that we might eventually have to find a way of making sense of non-renormalizable theories.

Salam and Strathdee proposed a way of dealing with theories with non-polynomial interactions. In earlier work, particularly by the Soviet physicists G.V. Efimov and E.S. Fradkin, it had been suggested that although interactions of finite degree greater than four lead to irremovable divergences, it could be that by including interactions of *all* orders one might find better behaviour, perhaps even finiteness without the need for *any* infinite renormalization. In their 1970 paper, Salam and Strathdee developed a momentum-space approach, based on the work of another Soviet physicist, M.K. Volkov. They devised a graphical technique, with 'superlines' each representing a Borel sum over the contributions of all numbers of lines from 0 to ∞.

Although this work did not, as they hoped, lead directly to viable theories of the strong interactions and of gravity, the techniques used are in many ways quite similar to those that were successfully deployed some years later in dealing with string theories (see Sec. 4 of these collected papers).

The final paper in this first section builds on the now classic paper of Sidney Coleman and Erick Weinberg, in which they suggest that spontaneous symmetry breaking in a massless scalar theory with $\lambda\phi^4$ interaction could be a purely quantum-mechanical effect. They discussed the renormalization of such a theory and showed how it could lead to the intriguing phenomenon of 'dimensional transmutation', in which the dimensionless parameter λ is 'traded' for a dimensionful parameter σ, the position of the minimum of the effective potential. Salam and Strathdee showed how this result could be obtained using the framework of the Dyson-Schwinger integro-differential equations, a result of very considerable importance.

Reprinted from the Maths. Student—Vol. XI. Nos. 1-2, Mar.-June 1943

A Problem of Ramanujam †

Solve

$$x^2 = a + y \quad \ldots \text{(i)}$$
$$y^2 = a + z \quad \ldots \text{(ii)}$$
$$z^2 = a + u \quad \ldots \text{(iii)}$$
$$u^2 = a + x \quad \ldots \text{(iv)}$$

1. Suppose x, y, z, u are the roots of a biquadratic

$$t^4 + p_1 t^3 + p_2 t^2 + p_3 t + p_4 = 0$$

We denote $\sum x^n$ by S_n.

Now $S_1 = -p_1$

$S_2 = 4a + S_1 = 4a - p_1$ from the given equations ... (a)

Also $S_2 + p_1 S_1 + 2p_2 = 0$

$\therefore$ Substituting for S_1, S_2 we have $p_2 = \dfrac{p_1^2 + p_1 - 4a}{2}$ ... (b)

2. Subtract (iii) from (i) and (iv) from (ii)

$x^2 - z^2 = y - u$; $y^2 - u^2 = z - x$

$\therefore$ $(x^2 - z^2)(y^2 - u^2) = (y - u)(z - x)$

or $xy + zy + ux + uz = -1$

But $\sum xy = p_2$

So $xz + uy = p_2 + 1$ ... (c)

3. $x^3 = ax + xy$

$y^3 = ay + yz$

Adding $S_3 = aS_1 + xy + zy + ux + uz$

$= -ap_1 - 1$ ... (d)

Also $x^2z^2 = a^2 + au + ay + uy$

$y^2u^2 = a^2 + az + ax + zx$

Adding, $x^2z^2 + y^2u^2 = 2a^2 + aS_1 + uy + zx$

$(xz + uy)^2 = 2a^2 + aS_1 + p_2 + 1 + 2p_4$ $[xyzu = p_4]$

$(p_2 + 1)^2 = 2a^2 - ap_1 + p_2 + 1 + 2p_4$

$\therefore$ $2p_4 = p_2^2 + p_2 - 2a^2 + ap_1$ ... (e)

4. Evidently

$(x^2 - y^2)(y^2 - z^2)(z^2 - u^2)(u^2 - x^2) = (x - y)(y - z)(z - u)(u - x)$

or $(x + y)(y + z)(z + u)(u + x) = 1$

or $(x^2z^2 + u^2y^2 + 2xyzu) + \sum x^2yz = 1$

$(p_2 + 1)^2 + p_1 p_3 - 4p_4 = 1$ $\{\sum x^2yz = p_1 p_3 - 4p_4\}$

Substituting for p_4 from (e)

$p_1 p_3 - p_2^2 + 4a^2 - 2ap_1 = 0$ ... (f)

† Vide *Collected Papers of Srinivasa Ramanujan*, 1927, p. 332 Q. 722 Also J. I. M. S. Series I Vol. VII p. 240.

5. We know

$$S_3 + p_1 S_2 + p_2 S_1 + 3p_3 = 0$$

Substituting for S_3, S_2, S_1, p_2

we get

$$6p_3 = 3p_1^2 + p_1^3 - 10\,ap_1 + 2 \qquad \ldots \quad (g)$$

Substitute for p_3 from (g) in (f)

we get

$$p_1^4 - (4a-3)\,p_1^2 - 4p_1 = 0 \qquad \ldots \quad (h)$$

$$p_1\,\{p_1^3 - (4a-3)\,p_1 - 4\} = 0$$

The cubic in p_1 can be solved by the usual methods. p_1 known, p_2, p_3, p_4 can easily be discovered and the corresponding biquadratic in t can be framed. The biquadratic may be solved by usual methods for x, y, zu; For the particular value $p_1 = 0$ the biquadratic is

$$t^4 - 2a\,t^2 + \tfrac{1}{3}\,t + a^2 - a = 0$$

6. By employing the same methods, we can solve the system of equations

$$x^3 = a + y$$
$$y^3 = a + z$$
$$z^3 = a + x$$

much more rapidly than Ramanujan did. His is a very laborious method.

Govt. College, Lahore.

ABDUS SALAM,
Fourth Year Student

St. Joseph's I. S. Press, Trichy—1944.

Reprinted from THE PHYSICAL REVIEW, Vol. 82, No. 2, 217-227, April 15, 1951

Overlapping Divergences and the S-Matrix

ABDUS SALAM
St. John's College, Cambridge, England
(Received September 29, 1950)

By extending considerations given by Dyson, general rules are obtained for isolating divergent parts from integrals corresponding to overlapping graphs, and a proof is obtained for the appearance of an extra factor Z_1^{-1} from "b divergences." In the last section the possibility of renormalization for scalar meson-nucleon interactions is demonstrated.

I. INTRODUCTION

IN his treatment of spinor electrodynamics Dyson[1] has defined operators Λ_μ, Σ^*, and Π^* corresponding to the three types of primitive divergent graphs in the theory. In the calculation of the contribution to Γ_μ arising from a reducible vertex part V_R it is possible to break V_R down unambiguously into an irreducible vertex part plus various inserted self-energy (S) and vertex (V) parts. The divergences introduced by the latter can be removed in a well-defined manner because any two of the insertions made in V_R are either completely non-overlapping or else are so arranged that one is completely contained in the other. This procedure fails, however, in the calculation of the contributions to Σ^* or Π^* from reducible self-energy graphs. Considering, for example, the electron self-energy, there is just one irreducible graph W_1 (Fig. 1). V parts inserted at one of the two end vertices a or b appear simultaneously as vertex insertions at the other vertex. Correspondingly, the contribution to Σ^* arising from a reducible part W_R is, in general, an integral which involves divergences corresponding to each of the ways in which W_R might have been built up by insertion of V parts at either or both vertices of W_1. Dyson has called these "b-divergences" and their expected effect is appearance of an extra factor Z_1^{-1} in his Eqs. (88) and (89). In order to demonstrate the possibility of renormalization, it is vital that this factor should appear; and it is the purpose of this paper to attempt a formal proof. The considerations presented here throw some light on the prospects of renormalization for scalar electrodynamics.

II. SEPARATION OF OVERLAPPING DIVERGENCES

Dyson[2] (unpublished) has defined a formal mathematical procedure for the isolation of the divergent part from an integral representing overlapping divergent graphs. This procedure is illustrated most readily by an example.

Figure 2 represents an electron self-energy graph which could be obtained by the insertion of a V part at vertex a or b of Fig. 1. Here

$$\Sigma(W_2, p)=e^4\int\int dt_1dt_2F_\mu(p, t_1)G_{\nu\mu}(p, t_1, t_2)H_\nu(p, t_2), \quad (1)$$

where

$$F_\mu(p, t_1)=\frac{1}{t_1^2}\gamma_\mu\frac{i\gamma(p-t_1)-\kappa}{(p-t_1)^2+\kappa^2},$$

$$G_{\nu\mu}(p, t_1, t_2)=\gamma_\nu\frac{i\gamma(p-t_1-t_2)-\kappa}{(p-t_1-t_2)^2+\kappa^2}\gamma_\mu,$$

$$H_\nu(p, t_2)=\frac{i\gamma(p-t_2)-\kappa}{(p-t_2)^2+\kappa^2}\gamma_\nu\frac{1}{t_2^2}.$$

Not only is the double integral over t_1t_2 linearly divergent, but it also diverges logarithmically if the integration is performed over either t_1 or t_2, while the other variable is held fixed. In order to isolate these divergences, we use (here and in the subsequent work) the invariant separation procedure outlined in Sec. VI of D II.

Rewrite Eq. (1) as follows:

$$\begin{aligned}\Sigma(W_2, p)=e^4\int\int dt_1dt_2[&F_\mu(p, t_1)G_{\nu\mu}(p, t_1, t_2)H_\nu(p, t_2)\\&-F_\mu(p_0, t_1)G_{\nu\mu}(p_0, t_1, 0)H_\nu(p, t_2)\\&-F_\mu(p, t_1)G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2)]\\&+e^4\left(\int dt_1F_\mu(p_0, t_1)G_{\nu\mu}(p_0, t_1, 0)\right)\\&\times\left(\int dt_2H_\nu(p, t_2)\right)+e^4\left(\int dt_1F_\mu(p, t_1)\right)\\&\times\left(\int dt_2G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2)\right), \quad (2)\end{aligned}$$

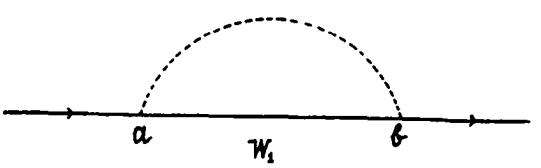


FIG. 1.

[1] F. J. Dyson, Phys. Rev. 75, 1736 (1949), referred to as D II, in this paper.

[2] The only published calculation for an integral corresponding to an overlapping (fourth order) graph is that of R. Jost and J. M. Luttinger, Helv. Phys. Acta 23, 201 (1949), who have followed Dyson's procedure.

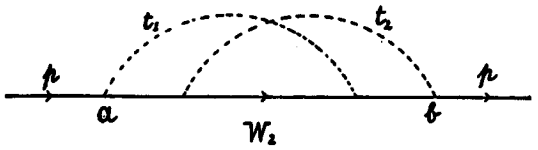


FIG. 2.

where $p_0^2+\kappa^2=0$, $i\gamma p_0+\kappa=0$. If

$$\int dt_1 F_\mu(p_0, t_1)G_{\nu\mu}(p_0, t_1, 0)=L_a\gamma_\nu,$$

$$\int dt_2 G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2)=L_b\gamma_\mu,$$

then we write

$$\Sigma(W_2, p)=e^4\int\int dt_1dt_2[R(p, t_1, t_2)]$$

$$+e^4L_a\int\gamma_\nu H_\nu(p, t_2)dt_2+e^4L_b\int F_\mu(p, t_1)\gamma_\mu dt_1, \quad (3)$$

where $R(p, t_1, t_2)$ is the expression within square brackets in Eq. (2).

The terms within the square bracket are such that if t_2 is held fixed, the integration over t_1 is convergent if the first and the second terms are combined—this second term being obtained by putting in those factors of $F_\mu(p, t_1)G_{\nu\mu}(p, t_1, t_2)H_\nu(p, t_2)$ which contain t_1, all momenta other than t_1 equal to their free particle values—while, if t_1 is held fixed, the first and the third terms in the square bracket combine to give a convergent t_2 integration.

Algebraically, it is impossible to find a single term which if subtracted might make the t_1 and t_2 integrations convergent simultaneously. Even as it stands, the expression within the square brackets, is, as a whole, convergent neither over t_1 nor over t_2, nor has this subtraction made any difference in its degree of divergence when the double integration is performed. However, when t_2 is held fixed and the integration over t_1 is performed, the third term in the square bracket

$$-\int\int dt_1dt_2F_\mu(p, t_1)G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2)$$

can be written

$$-\left\langle\int\int dt_1dt_2[F_\mu(p, t_1)-F_\mu(p_0, t_1)\right.$$

$$\left.-(p-p_0)(\partial/\partial p)F_\mu(p_0, t_1)](G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2))\right\rangle$$

$$-\left\langle\int dt_1(F_\mu(p_0, t_1)+(p-p_0)(\partial/\partial p)F_\mu(p_0, t_1))\right\rangle$$

$$\times\left\langle\int dt_2G_{\nu\mu}(p_0, 0, t_2)H_\nu(p_0, t_2)\right\rangle.$$

The integrand in the first angular bracket, $\langle\ \rangle$, above is absolutely convergent as far as the t_1 integration is concerned, while the second angular bracket, $\langle\ \rangle\langle\ \rangle$, is a product of a divergent constant L_b with a term of the form $A+B(\gamma p-i\kappa)$. Hence, with t_2 fixed, the integral of $R(p, t_1, t_2)$ over t_1 consists of an absolutely convergent part together with a divergent part, the latter being unambiguous and fully determinate. Since a similar result holds for t_2, we can say that the expression in the square bracket is "convergent" in a special sense over both the t_1 and t_2 integrations. It is to be emphasized that this has been made possible because, in the second (and also the third) term in the square bracket, the overlap factor G no longer appears as a function of two variables and so the integrand $F_\mu(p_0, t_1)G_{\nu\mu}(p_0, t_1, 0)H_\nu(p, t_2)$ splits into a product of two functions, each a function of a single variable.

In order to obtain the absolutely convergent (and physically significant) part of $\Sigma(W_2, p)$, we now perform the final separation. Thus,

$$\Sigma_c(W_2, p)=e^4\int\int dt_1dt_2[R(p, t_1, t_2)-R(p_0, t_1, t_2)$$

$$-(p-p_0)(\partial/\partial p)R(p_0, t_1, t_2)]. \quad (4)$$

If Eq. (4) is written out in full, we notice the basic point of this procedure. Whereas, in Eq. (4), we have subtracted terms with a view to securing convergence over the double integration, we have also succeeded, by the same step, in making $\Sigma_c(W_2, p)$ convergent for each of the two t_1 and t_2 integrations. The part which we designated above as the unambiguous and fully determinate divergent part of the t_1 integration, has in fact canceled out.[2a] The first subtraction in Eq. (2) made the expression "convergent" in our special sense over and t_2, the final subtraction in Eq. (4) made it absolutely convergent over all the t_1, t_2 and t_1t_2 integrations. With the final step we have succeeded in isolating from the integral, divergences corresponding to the inserted V parts as well as the divergence corresponding to the over-all S part.

III. GENERAL RULES FOR THE SEPARATION OF DIVERGENCES

The procedure of Sec. II for the separation of divergent parts can be generalized to apply to any divergent n-fold integral howsoever complicated by overlapping or non-overlapping insertions. Consider a graph having n basic[3] momentum-vectors t_i associated with internal lines, such that the momentum vectors of the remaining lines are expressed in terms of these n vectors. The contribution to the corresponding matrix element is an integral, I, over n variables.

In general the integral will consist of a product of functions of each variable t_i (and possibly of momenta corresponding to external lines) with overlap functions,

[2a] This has happened because the integrals multiplying L_a and L_b in Eq. (3) are themselves self-energy integrals.

[3] P. T. Matthews, Phil. Mag. 41, 185 (1950).

i.e., functions of two or more variables, t_i. If a set of variables $t_a t_b \cdots t_m$ is held fixed, the integration over the remaining variables will be called a "subintegration." To estimate the convergence of each subintegration we employ the considerations given in Sec. V of D II and merely count the powers of the relevant t in the numerator and the denominator of the integrand. An n-fold integral, I, will be absolutely convergent if, besides the final integration, the subintegrations over all of the smaller sets of variables t_i, $t_i t_j$, $t_i t_j t_k$, $\cdots$, are also absolutely convergent for all possible choices of the basic variables. If by merely counting the powers of t_i we find that an integral is itself convergent, while any one of the subintegrations is not, the integral will be called "superficially convergent."

In general in an n-fold integral over $t_1 t_2 \cdots t_n$ a subintegration over $t_1 t_2 \cdots t_r$ can be made convergent only if we subtract from the integrand terms in which those factors in the integrand which are functions of $t_{r+1}, \cdots t_n$ only are left unchanged, while in the remaining factors the external momenta (which now include $t_{r+1}, \cdots t_n$) are given their free-particle values. Thus, to fulfill the condition of convergence over all subintegrations, we must subtract from the integral divergent terms corresponding to each subintegration, the degree of divergence of the integrand for a particular subintegration, in general, being unaffected by subtractions corresponding to other subintegrations.

The considerations of Sec. II can now be generalized. We define true divergence over a subintegration $t_1 t_2 \cdots t_r$ as the divergent part finally to be subtracted corresponding to this subintegration, in order to make the subintegration absolutely convergent, after the terms involving divergences corresponding to t_i, $t_i t_j$, $t_i t_j t_k$, $\cdots$, $t_1 t_2 \cdots t_{r-1}$, $\cdots$ subintegrations have already been separated. Anticipating the procedure to be formulated, we shall find that the above-mentioned divergences, corresponding to these latter integrations over smaller number of variables are themselves their true divergences. Thus, for example, to obtain the true divergence over $t_1 t_2$ subintegration we hold, first, all variables except t_1 fixed and subtract from the integrand divergent terms corresponding to t_1, leaving the rest of the integrand unchanged; similarly, holding all variables except t_2 fixed, we subtract divergent terms corresponding to t_2. Finally, holding all variables except $t_1 t_2$ fixed in this *new* integrand, we subtract divergent terms corresponding to the double integration. These last constitute the true divergence over $t_1 t_2$ subintegration. The integrand after the separation of all these divergent parts is absolutely convergent as far as $t_1 t_2$ subintegration is concerned.

We can now give, in analogy with Sec. II, unambiguous rules for the separation of divergences from an n-fold integral. Before making the final subtraction corresponding to the n-fold integration which will make the remaining integral an absolutely convergent one, we subtract from the integrand divergent terms corresponding to each of the subintegrations t_i, $t_i t_j$, $t_i t_j t_k$, $\cdots$, $t_1 t_2 \cdots t_{n-1}$. The divergent term to be separated corresponding to the $t_i t_j$ subintegration would, for example, consist of an n-fold integral over $t_1 t_2 \cdots t_n$, in which the part of the integrand containing t_i or t_j is such as to give, on integration, the true divergence corresponding to $t_i t_j$ multiplied by those factors in the original integrand which contain neither t_i nor t_j and have thus been left unchanged. In this expression, the integration over all variables other than $t_i t_j$ will correspond to a graph which we shall call the "reduced graph," and the corresponding integral the "reduced integral." Thus, in Sec. II, $\int dt_2 \gamma_\nu H_\nu(p, t_2)$ and $\int dt_1 F_\mu(p, t_1)\gamma_\mu$ are reduced integrals.

Our final result can be stated thus. The entire divergence to be separated from an n-fold integral is equal to the true divergence over the t_1 integration multiplied by the reduced integral over $t_2 \cdots t_n$ plus the true divergence over the t_2 integration multiplied by the reduced integral over $t_1 t_3 \cdots t_n$ and so on, together with the true divergence over the $t_1 t_2$ integration multiplied by the reduced integral over $t_3 t_4 \cdots t_n$ and so on, together with similar sets of terms ending finally with the true divergence over the final $t_1 t_2 \cdots t_n$ integration. After all these divergent terms have been isolated, the remainder is an absolutely convergent integral.

The maximum possible number of terms which may need to be isolated is $2^n - 1$. Of course, not all of the subintegrations will be divergent, a considerable number being "superficially" convergent, so that no corresponding separations will need to be made.

It may be possible (and indeed as we shall see later, it is possible in some special cases) to secure the actual convergence of certain subintegrations before the true divergence of the n-fold integration is separated. However, that this last separation will always leave behind an absolutely convergent integral as remainder, would follow from a generalization of considerations given in Sec. II.[3a]

A few remarks on superficial convergence will be relevant at this stage.

(a) In spinor electrodynamics all graphs with four or more external lines (except graphs corresponding to scattering of light by light) are at least superficially convergent. The "scattering of light by light" itself is not a genuine primitive divergent on account of the gauge invariance of the theory. It should, therefore, always be possible to find other graphs which when combined with the graph in question should give a convergent result. We shall therefore treat the integrals corresponding to the "scattering of light by light" too, as being at least formally superficially convergent. Opening two or more lines in a connected self-energy graph always leads to superficial convergence in the connected[3b] part of the remainder and so does the

[3a] A general proof of this will be published elsewhere.

[3b] Any part of a connected graph is joined to the rest by at least two lines.

opening of one or more lines in a vertex part, unless, of course, so many lines are opened that the only connected graph left is a V part. These considerations will immediately give the number of superficially convergent integrations, when dealing with S or V parts.

If an internal line l in a graph contains a self-energy insertion with associated momenta $l_1, l_2, \cdots l_r$, then the subintegrations over ll_i, ll_il_j, $\cdots$, $ll_1l_2\cdots l_{r-1}$, $\cdots$ (but not $ll_1l_2\cdots l_r$) are superficially convergent. l, being an internal line, itself belongs to a loop, which at the very worst consists of two electron lines, with but two external photon lines, the whole forming as far as $ll_1\cdots l_r$ integration is concerned, a photon self-energy graph. Opening one single line belonging to the self-energy insertion inside l gives at worst a scattering of light by light graph. Thus the subintegrations ll_i, ll_il_j, $\cdots$, $ll_1\cdots l_{r-2}$, $\cdots$ are certainly superficially convergent, while with the convention adopted above for the scattering of light by light graphs, the subintegrations over the last set $ll_1l_2\cdots l_{r-1}$, $\cdots$ are also superficially convergent.

(b) If two sets of variables $(l_i\cdots l_j)$ and $(l_a\cdots l_b)$ are such that the corresponding graphs do not overlap and if the integration over each set has been made convergent, then the product of these convergent integrands is also convergent over the subintegrations l_i, l_a, l_il_a, $\cdots$, $l_i\cdots l_jl_a\cdots l_b$. For integration of functions with no overlap terms certain new features arise. Taking a simple example, the integral $\int\int dl_1dl_2F(p, l_1)G(p, l_2)$ has no overlap term, and thus, if both l_1 and l_2 subintegrations are divergent (in this example logarithmically),

$$\int\int dl_1dl_2[F(p, l_1)-F(p_0, l_1)][G(p, l_2)-G(p_0, l_2)]$$

is absolutely convergent. We notice that the divergent part separated is

$$\int\int dl_1dl_2[F(p, l_1)G(p_0, l_2) + F(p_0, l_1)G(p, l_2)-F(p_0, l_1)G(p_0, l_2)].$$

Although it loses its precise significance, we may still formally retain the expression "true divergence" over l_1 for $\int dl_1F(p_0, l_1)$, over l_2 for $\int dl_2G(p_0, l_2)$, and over l_1l_2 for $-\int\int dl_1dl_2F(p_0, l_1)G(p_0, l_2)$. The negative sign before the true divergence from l_1l_2 integration, when overlap does not occur, makes the use of our general rule inconvenient for non-overlapping graphs.

The above remarks do show, however, that for non-overlapping graphs it is immaterial in what order divergences are removed, while for divergences arising from graphs one of which is completely contained in the other, Dyson's prescription of successive removal starting from the innermost insertion is a particular case of the general rule given above. We shall illustrate this rather more fully by considering self-energy insertions.

(c) Suppose a line l, in a graph I, contains a self-energy insertion with momenta $l_1, l_2, \cdots l_r$. It will be convenient to remove divergent parts from the integral corresponding to subintegrations $l_1, \cdots l_1l_2, \cdots, l_1l_2\cdots l_r$. After the "true divergence" over $l_1l_2\cdots l_r$ has been removed, the subintegration over this set is left absolutely convergent. The momenta associated with the self-energy insertion in line l define (except for l), no overlap function with any other momentum variable l_a in I. According to Remark (b), if the integrand is made convergent over l_a, its convergence over l_al_1, $\cdots l_al_1l_2\cdots l_r$, is already assured. Remark (a) shows that the same result holds for the only variable (l) which does have overlap functions with these variables. At this stage the only integrations in which $l_1, l_2, \cdots l_r$ appear explicitly are the $ll_1l_2\cdots l_r$ subintegration, and other subintegrations in which these variables appear as a group. But from considerations given by Dyson we know already that, for questions of convergence, this group will behave precisely as though the line l had no insertion inside it. This shows that it is immaterial whether we retain the variables $l_1, l_2, \cdots l_r$ or perform the integration over them at this stage; the latter procedure implies replacing $S_F(l)$ by $S_c(l)S_F(l)/2\pi$, which is precisely the result one would get if one followed the procedure of successive removal as described in D II. By using results on superficially convergent subintegrations derived by applying Remark (a), a similar proof could be constructed to show the validity of the procedure of successive removal for divergence arising from V parts one of which is completely contained in the other.

IV. APPLICATION TO ELECTRON SELF-ENERGY GRAPHS

We shall now apply the above considerations to the case of electron self-energy graphs. According to the Dyson-Feynman procedure for obtaining the effective radiative corrections to any physical process, one draws the relevant irreducible graphs and then replaces S_F by S_{F1}', D_F by D_{F1}', and γ_μ by $\Gamma_{\mu 1}$ in the corresponding integrals. One already uses the same procedure for calculating $\Gamma_{\mu 1}$ itself (Sec. VII, D II); one draws all of the irreducible vertex graphs and then makes the above substitutions. We attempt to obtain a similar set of self-energy graphs in which these substitutions can be made unambiguously—only in this case this set of graphs will not be irreducible. For this purpose we establish the following categories of self-energy graphs.

Category [1] consists of the sole irreducible self-energy graph in Fig. 1.

Category [2] consists of all graphs formed by inserting[4] at the end vertex b in Fig. 1 all irreducible V parts of any order in e. The first end vertex of each category $[n]$ will be called a and the last b, although sometimes, for clarity, we use $a_{[n]}$ and $b_{[n]}$.

To obtain the graphs in category [3] insert at $b_{[2]}$ all

[4] End vertices of self-energy graphs, vertex graphs, or Compton graphs, are the vertices where the external electron line enters or leaves the graph.

irreducible vertex parts, and so on for all subsequent categories.

Continuing the above procedure we obtain an infinite number of categories, each (except [1]) containing an infinite number of graphs. Figure 2, for example, belongs to category [2]. All graphs in $[n]$ can, in fact, be built up in precisely n different ways from the graph in Fig. 1 by the insertion of vertex parts at $a_{[1]}$ or $b_{[1]}$ or both. Further, all graphs in $[n]$ could equally well have been built by insertions at $a_{[n-1]}$ rather than at $b_{[n-1]}$. This complete symmetry will be found to be important for the subsequent proof.

A graph belonging to $[n]$ will contain precisely $(n-2)$ photon lines such that if these are opened (i.e., their associated integration variables held fixed), the graph splits into two vertex parts. Given a graph in $[n]$ we denote by L_b^1 the divergent constant arising from the irreducible V part which was inserted at $b_{[n-1]}$ in order to obtain this particular graph in $[n]$, by L_b^2 the "true" divergent constant arising from the reducible part which was inserted at $b_{[n-2]}$ to obtain this graph in $[n]$, and so on. This last V part is reducible because its end vertex has as an insertion the V part with the divergent constant L_b^1. Similar definitions apply for $L_a^1, L_a^2, \cdots$. We shall now prove the following lemmas.

Lemma 1.—Denote by $[n]$ the integral corresponding to a particular graph in the category $[n]$. We have the following result.

$$[n]=(L_a^1[n-1]+L_a^2[n-2]+L_a^3[n-3]+\cdots+L_a^{n-1}[1])+(L_b^1[n-1]+L_b^2[n-2]+L_b^3[n-3]+\cdots+L_b^{n-1}[1])-\left\langle \begin{matrix} L_a^1L_b^1[n-2]+L_a^1L_b^2[n-3]+L_a^1L_b^3[n-4]+\cdots \\ +L_a^2L_b^1[n-3]+L_a^2L_b^2[n-4]+\cdots \end{matrix} \right\rangle +\Sigma_{[n]}^*, \quad (5)$$

where

$$\Sigma_{[n]}^*=A_{[n]}+B_{[n]}(\gamma p-i\kappa)+\Sigma_{a[n]}^*.$$

A and B characterize the true divergence from $[n]$, and $[n-r]$ stands for an integral corresponding to a certain graph in category $[n-r]$ obtained from $[n]$ by successive removal of V parts from a or b as the case may be.

Proof.—It is convenient to associate the basic variables with the photon lines in the graph. Suppose $[n]$ has k basic variables of which $n-2$ are critical. The photon line t_1 starts from $a_{[n]}$, while t_k ends at $b_{[n]}$.

For the proof we shall systematize our procedure as follows: we hold each one of the variables t_i initially fixed, in turn, and remove all of the divergent terms corresponding to all possible subintegrations over the remaining $(n-1)$ variables. Of course, in following this procedure we shall have to guard against subtracting a divergent term more than once. However, holding t_k fixed we make the removal unambiguously corresponding to the subintegrations $t_1, \cdots, t_{k-1}, t_1t_2, \cdots, t_1t_{k-1}, t_2t_3, \cdots, t_2t_{k-1}, \cdots, t_1t_2t_3, \cdots, t_1t_2\cdots t_{k-1}$. This immediately gives us the terms in the first bracket of Eq. (5) by our general rule. The constants L_a^1, L_a^2, $\cdots$, etc., come from the successive reduction of a reducible vertex parts. It is to be emphasized that the first line in Eq. (5) is the result of the removal of divergences, corresponding to all the above subintegrations, due regard being paid to those subintegrations which are superficially convergent [see Remark (a)]. Now holding t_1 fixed we remove divergent terms corresponding to all the possible subintegrations over subsets of variables $t_2, t_3, \cdots t_k$. A number of these terms (those, for example, corresponding to subintegrations over $t_2, t_3, \cdots t_{k-1}, t_2t_3, \cdots t_2t_{k-1}, \cdots$, etc.) have already been removed. However, the subintegrations from these subsets which still remain, i.e., over $t_k, t_2t_{k-1}, \cdots t_2t_3, \cdots t_2t_3t_{k-1}t_{k-2}, \cdots$ (barring those which are superficially convergent), cover the successive divergent parts of the (reducible) vertex graph one obtains by opening the line t_1 and are just sufficient to give us the second bracket in Eq. (5). The maximum number of subintegrations corresponding to which divergent terms still remain to be separated is 2^{k-2} at this stage. These terms must be separated by holding each of the remaining $k-2$ basic variables initially fixed. Of these only $(n-2)$ are critical. As noted earlier, opening a critical line splits the graph into two disconnected segments with no overlap term between them. As an example, let us assume t_2 to be the first critical variable, occurring from the left. Holding it fixed, we notice that the only divergent terms yet to be separated are the true divergences corresponding to the subintegrations $t_1t_2, t_1t_2t_{k-1}, t_1t_2t_3t_{k-2}, \cdots$, those corresponding to $t_1, t_k, t_2t_{k-1}, t_2t_3t_{k-2}, \cdots$, etc., already having been removed. Since these are the true divergences for integrations over non-overlapping parts, we shall get a negative sign before the separated terms as noted in Remark (b). This leads to the isolation of divergent terms of the form $-L_a^1L_b^1[n-2]$. Starting from end a and opening all the critical lines, we shall get all the subtracted terms on the right-hand side of the Eq. (5). It remains to prove now that holding each one of the remaining $(k-n)$ lines fixed gives no further divergent terms to be separated. We notice that every other line, if opened, converts the graph into a C part,[5] which we know to be at least superficially convergent in spinor electrodynamics. If the C part obtained by opening a line is reducible, the only insertions its irreducible skeleton has are vertex parts, inserted at its end vertices. But the divergent terms corresponding to the subintegrations over the variables associated with these vertex insertions have already been removed by the successive removal performed by holding t_1 and t_2 fixed initially. Thus, these C parts give no further contribution to Eq. (5). This establishes the lemma.

If we had started with basic variables associated with electron rather than with photon lines, opening a basic line would convert the graph into an M part[6]—again at least superficially convergent—and similar considerations would apply.

Lemma 2.—This is a restatement of Remark (c) for the particular case of graphs in these categories. The divergent separations after inserting S parts in all the lines or V parts in all the *internal* vertices (all vertices except a and b) of $[n]$ are given by the divergent parts corresponding to the subintegrations over the internal variables of the insertions [as in Remark (c)], together with the divergent terms to be separated from $[n]$, where $S_F(t)$ is replaced by $S_c(t)S_F(t)$, $D_F(t)$ by $D_F(t)D_c(t)$, and γ_μ by $\Lambda_{\mu\epsilon}(t, t')$. As a concrete example, we see that if a photon self-energy graph is inserted in a line t of $[n]$, such that $D_F(t)\Pi^*(t)=C+D_c(t)$, then the resulting integral is equal to $C[n]+[n]^\times$, where $[n]^\times$ is the integral obtained by replacing $D_F(t)$ by $D_c(t)D_F(t)$ in the integrand. The proof is exactly similar to that of Remark (c).

If we now desire to construct all self-energy graphs up to a given order e^{2k+2}, say, we draw all of the graphs in the various categories defined above up to this order. No graph belonging to category $[k+2]$ or higher will be needed, the lowest order graph belonging to the category $[k+2]$ being of the order e^{2k+4}. Taking each one of the graphs drawn above, V and S parts will be inserted in its internal vertices and all its lines. *No insertions are ever to be made in the end vertices of the graphs of any category.* If the graph in which insertions are being made is of order $2r$, these insertions need only consist of graphs up to the order $(2k-2r+2)$.

It is easy to verify now that the above procedure gives accurately all self-energy graphs up to order $(2k+2)$ and that no graph appears more than once. Among them we shall, of course, find all the graphs which could have been obtained by insertions of vertex parts up to order $2k$ at the vertices a (or b) of the fundamental self-energy graph in Fig. 1. To verify the appearance of all vertex parts, up to the given order, we notice, for example, that by the very principle of their construction the graphs of

[5] A Compton graph—a graph with 2 external photon and 2 external electron lines.

[6] Møller graph—a graph with four external electron lines.

category [3] represent the insertion of vertex parts in the end vertices of graphs of category [2], while the insertions in their own end vertices are given by the graphs of the higher categories.

Correspondingly, adopting the procedure in Sec. VII of D II of dropping the divergent constants when they appear, if we assume that S_{F1}', D_{F1}', $\Gamma_{\mu 1}$ are defined up to order e^{2k} by, respectively, S_F plus S_F multiplied by a finite sum of products of $S(W,t)$, D_F plus D_F multiplied by a finite sum of products of $D(W,t)$, and γ_μ plus a finite sum of $\Lambda_{\mu\epsilon}(t,t')$, we can obtain S_{F1}' up to the order e^{2k+2} by drawing all of the graphs in the various categories, up to the order e^{2k+2}, and then substituting in all lines S_{F1}', D_{F1}' *(defined as above)* instead of S_F and D_F, and $\Gamma_{\mu 1}$ in all internal vertices instead of γ_μ. The γ_μ occurring at the two end vertices are always to be left unchanged. Once again, in a graph of order $2r$ belonging to any one of the categories, it will only be necessary to substitute S_{F1}', etc., up to the order $(2k-2r+2)$.

Lemma 3.—If by the above changes in all of the lines and internal vertices of an integral arising from a graph in category $[n]$ we obtain an integral $[n]^\times$, then

$$[n]^\times = L_a{}^{1\times}[n-1]^\times + L_a{}^{2\times}[n-2]^\times + L_a{}^{3\times}[n-3]^\times + \cdots + L_b{}^{1\times}[n-1]^\times + L_b{}^{2\times}[n-2]^\times + L_b{}^{3\times}[n-3]^\times + \cdots - (L_a{}^{1\times}L_b{}^{1\times}[n-1]^\times + L_a{}^{1\times}L_b{}^{2\times}[n-3]^\times + \cdots) + \Sigma^*{}_{[n]}{}^\times,$$

where $L_a{}^{1\times}$ stands for the "true" divergent constant from the V part introduced at $a_{[n-1]}$ in $[n-1]^\times$ in order to obtain $[n]^\times$. The definition of the other constants is similar. Noticing that $S_{F1}'(t)$ and $D_{F1}'(t)$, and $\Gamma_{\mu 1}$, as defined above, behave precisely as do $S_F(t)$, $D_F(t)$, and γ_μ as far as the power of variables involved and the question of convergence are concerned, the proof follows from Lemma 1.

Lemma 3 immediately gives the following set of equations

$$\begin{aligned}
&[1]^\times = \Sigma^*{}_{[1]}{}^\times \\
&[2]^\times = (L_a{}^{1\times} + L_b{}^{1\times})[1]^\times + \Sigma^*{}_{[2]}{}^\times \\
&[3]^\times = (L_a{}^{1\times} + L_b{}^{1\times})[2]^\times + (L_a{}^{2\times} + L_b{}^{2\times})[1]^\times \\
&\qquad\qquad - L_a{}^{1\times}L_b{}^{1\times}[1]^\times + \Sigma^*{}_{[3]}{}^\times \\
&\quad\cdots\cdots\cdots

wave functions, that such graphs are in fact not divergent.[7]

In the sequel we shall consider pseudoscalar (charged or neutral) and scalar (charged) theories. Thus, besides the primitive divergents in electrodynamics the only new feature would be the divergent graphs corresponding to the Møller scattering of two mesons (M-parts). Following a suggestion first given by Matthews, it will be proved that the divergences arising from these M parts can be canceled consistently by adding an appropriate δ-type interaction between mesons to the lagrangian. We consider in detail only the case of pseudoscalar neutral meson, the remaining cases being almost identical.

Following Matthews we write the interaction hamiltonian as

$$H_1(x)=if\bar{\psi}(x)\gamma_5\psi(x)\phi(x)-\hbar c\delta\kappa_0\bar{\psi}(x)\psi(x)-\tfrac{1}{2}\delta\kappa^2\phi^2(x)-\delta\lambda\phi^4(x).$$

The corresponding graphs will have, besides the vertices with two nucleon and one meson line incident, also vertices with two nucleon, two meson, or four meson lines incident. These last will be called 4-vertices, while graphs formed entirely from 3-vertices will be called the "original graphs."

The following remark will apply to all except meson self-energy graphs, which will be considered later.

Corresponding to any original graph (formed entirely from 3-vertices), which itself represents a Møller scattering of mesons, or contains M parts inside itself, there exist in the theory graphs formed by replacing the M parts by 4-vertices *in all possible ways*. It is easy to verify that for a field theory with ϕ^4 as the only interaction term there are two primitive divergent graphs, graphs with two external meson lines (meson self-energy graphs), and graphs with four external meson lines (M parts).

We shall call a graph "simple" if it does not contain an M part inside itself, or is a Møller graph which cannot be formed by joining two or more M parts. We shall retain the definition of reducibility of a graph as adopted in electrodynamics. Thus, a graph may not be "simple" but may still be irreducible if it satisfies the criteria of irreducibility as in electrodynamics.

To choose the constant $\delta\lambda$ we proceed in three steps.

(1) Consider the category of all original M parts which are both "simple" and irreducible. If $\delta\lambda$ is chosen

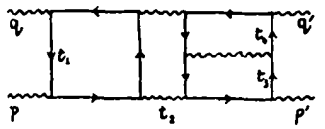


FIG. 3.

equal to the sum of their true divergences (the true divergence of an M part consists of a constant), the term $-\delta\lambda\phi^4$ suffices to cancel all divergences arising from the above graphs. $\delta\lambda$ is thus a power series in f.

(2) Now consider the category of all original irreducible M parts which are not simple. Before removing the true divergence from the over-all M part, divergent terms have to be removed corresponding to the M parts inside the graph in accordance with the procedure given in Sec. III. We illustrate the points involved with a simple example. Figure 3 represents an (irreducible) M part obtained by joining two simple M parts. The corresponding integral is given by[8]

$$\begin{aligned}f^{10}\int dt_1dt_2dt_3dt_4[&\text{Trace }\gamma_5S(p+t_1)\gamma_5S(p+t_1-t_2)\\&\times\gamma_5S(t_1-q)\gamma_5S(t_1)]\Delta(t_2)\Delta(p+q-t_2)\\&\times[\text{Trace }\gamma_5S(p'+t_3)\gamma_5S(t_3)\gamma_5S(t_4)\gamma_5S(t_4-q')\\&\times\gamma_5S(p'+t_4-t_2)\gamma_5S(p'+t_3-t_2)]\Delta(t_4-t_3).\end{aligned}$$

The integral is logarithmically divergent with respect to the t_1, t_2t_4, and $t_1t_3t_4$ subintegrations, besides being logarithmically divergent with respect to the final 4-fold integration. The remaining subintegrations are at least superficially convergent.

The divergent terms to be separated before the true divergence over $t_1t_2t_3t_4$ is separated, in order to leave behind an absolutely convergent integral are

$$\begin{aligned}(\delta\lambda)_1f^8\int dt_2dt_3dt_4\Delta(t_2)\Delta(p+q-t_2)&\\\times[\text{Trace }\gamma_5S(p'+t_3)\gamma_5S(t_3)\gamma_5S(t_4)\gamma_5S(t_4-q')&\\\times\gamma_5S(p'+t_4-t_2)\gamma_5S(p'+t_3-t_2)]\Delta(t_4-t_3)&\\+(\delta\lambda)_2f^4\int dt_1dt_2[\text{Trace }\gamma_5S(p+t_1)\gamma_5S(p+t_1-t_2)&\\\times\gamma_5S(t_1-q)\gamma_5S(t_1)]\Delta(t_2)\Delta(p+q-t_2)&\\-(\delta\lambda)_1(\delta\lambda)_2\int dt_2\Delta(t_2)\Delta(p+q-t_2),&\end{aligned}$$

where

$$\begin{aligned}(\delta\lambda)_1=f^6\int dt_1[&\text{Trace }\gamma_5S(p_0+t_1)\gamma_5S(p_0+t_1-t_{20})\\&\times\gamma_5S(t_1-q_0)\gamma_5S(t_1)]\\(\delta\lambda)_2=f^6\int dt_3dt_4[&\text{Trace }\gamma_5S(p_0'+t_3)\gamma_5S(t_3)\\&\times\gamma_5S(t_4)\gamma_5S(t_4-q_0')\gamma_5S(p_0'+t_4-t_{20})\\&\times\gamma_5S(p_0'+t_3-t_{20})]\Delta(t_4-t_3)\end{aligned}$$

and

$$p_0=q_0=p_0'=q_0'=t_{20},\quad p_0^2+\kappa^2=0$$

[7] *Note added in proof.*—Dyson has, in fact, proved that the contribution $F(p_1, p_2, p_3)$ from any graph with just three meson lines (p_1, p_2, p_3) in neutral pseudoscalar theory vanishes identically. Let p be a vector obtained from p by space reflection. Since $\phi(p_1)\phi(p_2)\phi(p_3)F(p_1, p_2, p_3)$ is a scalar, $F(p_1, p_2, p_3)$ is a pseudoscalar function and therefore $F(p)=-F(p)$. (This can be demonstrated by a more explicit mathematical proof.) But $F(p)$ is invariant under a proper Lorentz transformation; for example, under a space rotation, through 180° about an axis in line perpendicular to p_1 and p_2, which transforms p_1 to p_1 and p_2 to p_2, and since $p_1+p_2+p_3=0$, also transforms p_3 to p_3. Therefore, $F(p)=F(p)$. Hence, $F(p)$ must vanish identically.

The author is indebted to Mr. F. J. Dyson for kindly communicating the proof.

[8] S and Δ stand for Feynman's functions.

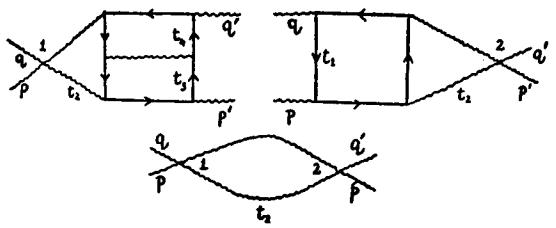


FIG. 4.

$(\delta\lambda)_1$, is the true divergence over t_1, $(\delta\lambda)_2$ over t_3t_4, and $-(\delta\lambda)_1\times(\delta\lambda)_2$ over $t_1t_3t_4$. The negative sign before the latter follows from the considerations about non-overlapping graphs given in Remarks in Sec. III.

Now we notice that on account of the presence of $-\delta\lambda\phi^4$ term in the hamiltonian with $\delta\lambda$ (for the present) chosen as the sum of true divergences from all irreducible simple M parts, there are three other graphs of order f^{10} in the theory, *viz.*, the graphs in Fig. 4. Since $\delta\lambda$ contains among other terms $(\delta\lambda)_1$ and $(\delta\lambda)_2$, we may assume that in the graphs drawn in Fig. 4 $-(\delta\lambda)_1\phi^4$ acts at the 4-vertex marked 1, and $-(\delta\lambda)_2\phi^4$ at the 4-vertex marked 2. Quite obviously the sum of contributions from these four graphs suffices to cancel all the divergent terms exhibited so far. The true divergence from the $t_1t_2t_3t_4$ integration $(\delta\lambda)_3$ which still remains to be separated from (1) to leave behind a convergent integral will now be added $\delta\lambda$ to already formed and we secure the cancellation of all divergences if for this process the graph consisting of a single 4-vertex $-(\delta\lambda)_3\phi^4$ is also taken into account. Thus if we consider all (irreducible) M parts which are not simple, only the true divergence over the final integration need be added to the $\delta\lambda$ term formed by adding the true divergences from all irreducible simple M parts. The result is of general validity and holds for the general case of an M part formed by joining any number of simple M parts.

It may be noticed that the operation of taking the trace does not interfere with the separation of divergences corresponding to an M part which is contained inside another. The reason is that nucleon lines run in loops while (in the case of original graphs) meson lines do not run continuously, thus making it possible to localize M parts unambiguously in this theory.

(3) $\delta\lambda$ so far has been obtained as a power series in f, and is equal to the sum of true divergences from all irreducible M parts. Considering reducible graphs now,

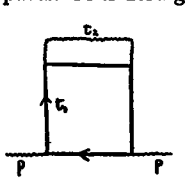


FIG. 5.

we have to consider the effect of inserting S and V parts in all the irreducible M parts. Since $S_F'=Z_2S_{F1}'(f_1)$, $\Delta_F'=Z_3\Delta_{F1}'(f_1)$, and $\Gamma^5=Z_1^{-1}\Gamma_1^5(f_1)$, we replace $S_F(p)$ by $S_{F1}'(p)$, $\Delta_F(p)$ by $\Delta_{F1}'(p)$, and γ_5 by $\Gamma_1^5(p,p')$ in all the lines and the vertices of the original irreducible M parts. An irreducible M part of order $2r$ contains $2r$ 3-vertices, $2r$ nucleon lines, and $r-2$ meson lines. The above substitutions give $f^{2r}Z_1^{-2r}Z_2^{2r}Z_3^{r-2}=Z_3^{-2}f_1^{2r}$ as the f-factor outside the integral corresponding to this irreducible M part, while the integral itself can be written as the sum of divergent and convergent terms, both expressed as power series in f_1. Extending the arguments given above for irreducible graphs we readily see that finally $\delta\lambda(f_1)=Z_3^{-2}\times M_d(f_1)$, where $M_d(f_1)$ is a power series in f_1 (starting with a term f_1^4) each of whose terms corresponds to the true divergence of some original M part.

So far we have proved that with the above choice of $\delta\lambda$ not only can all divergent terms be separated from the integral corresponding to the original M parts to be canceled, but further that the new M graphs introduced into the theory because of the term $-\delta\lambda\phi^4$ in the hamiltonian are also accounted for.

Similar arguments prove that if we are considering graphs corresponding to any other process (except meson self-energy graphs) which contain M parts inside them, a combination of the original graphs for the process, with graphs obtained by replacing the M parts by 4-vertices *in all possible ways* will always lead to a cancellation of all divergences arising from the contained M parts, in the original graphs. The rest of the proof for the finiteness of the S-matrix follows D II.

(B) Meson Self-Energy Graphs

We now consider the case of meson self-energy graphs. By joining up an incoming meson line to an outgoing meson line in any M part we obtain a meson self-energy graph. Conversely, in a meson self-energy graph, if any meson line is opened, we obtain an M part. Just as it was impossible to make the substitution $\Gamma_{\mu 1}$ for γ_μ at the end vertices of a self-energy graph in electrodynamics unambiguously, similarly, it is impossible to make an unambiguous substitution Δ_{F1}' for Δ_F (or S_{F1}' for S_F) if we desire to obtain a higher order meson self-energy graph from one of lower order.

However, the procedure for isolating divergences from M parts, outlined in Sec. III, leads to an essential simplification. The principles involved can be illustrated once again by considering a simple example.

Figure 5 represents a meson self-energy graph, in one of the nucleon lines of which a nucleon self-energy graph is inserted. The corresponding integral is

$$\Pi(p)=f^4\int dt_1dt_2[\text{Trace }\gamma_5S(t_1-p)\gamma_5S(t_1)\times\gamma_5S(t_1-t_2)\gamma_5S(t_1)]\Delta(t_2).$$

The integral is logarithmically divergent for the t_1 subintegration, linearly divergent for the t_2 subintegration, and quadratically divergent for the t_1t_2 integration. The divergent terms to be separated in order to leave

behind an absolutely convergent integral are the following:

$$f^4\int dt_1dt_2\,\mathrm{Trace}\,[\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\gamma_5S(t_1-t_{20})\gamma_5S(t_1)]\Delta(t_2)+f^4\int dt_1dt_2$$
$$\times\mathrm{Trace}\,[\gamma_5S(t_1-p)\gamma_5S(t_1)\{\gamma_5S(t_{10}-t_2)\gamma_5$$
$$+\langle(t_1-t_{10})\partial/\partial t_1\rangle_{t_1=t_{10}}(\gamma_5S(t_1-t_2)\gamma_5)\}S(t_1)\Delta(t_2)]$$
$$+\Big\{\Big\{f^4\int dt_1dt_2\,\mathrm{Trace}\,[\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\gamma_5S(t_1-t_2)\gamma_5S(t_1)\Delta(t_2)-\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\gamma_5S(t_1-t_{20})\gamma_5S(t_1)\Delta(t_2)-\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\{\gamma_5S(t_{10}-t_2)\gamma_5+\langle(t_1-t_{10})\partial/\partial t_1\rangle_{t_1=t_{10}}$$
$$\times(\gamma_5S(t_1-t_2)\gamma_5)\}S(t_1)\Delta(t_2)]\Big\}\Big\}$$
$$+f^4\int dt_1dt_2\{(p-p_0)_\mu(\partial/\partial p_\mu)$$
$$+\tfrac{1}{2}(p-p_0)_\mu(p-p_0)_\nu(\partial/\partial p_\mu)(\partial/\partial p_\nu)\}^{9}{}_{p=p_0}$$
$$\times\mathrm{Trace}\,[\gamma_5S(t_1-p)\gamma_5S(t_1)\gamma_5S(t_1-t_2)$$
$$\times\gamma_5S(t_1)\Delta(t_2)-\gamma_5S(t_1-p)\gamma_5S(t_1)$$
$$\times\{\gamma_5S(t_{10}-t_2)\gamma_5+\langle(t_1-t_{10})\partial/\partial t_1\rangle_{t_1=t_{10}}$$
$$\times(\gamma_5S(t_1-t_2)\gamma_5)\}S(t_1)\Delta(t_2)]. \quad (10)$$

where

$$i\gamma t_{10}+K_0=t_{10}^2+K_0^2=0$$

and

$$t_{20}^2+K^2=p_0^2+K^2=0.$$

Let

$$f^4\int dt_1\,\mathrm{Trace}\,[\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\gamma_5S(t_1-t_{20})\gamma_5S(t_1)]=\delta\lambda,$$
$$f^2\int dt_2,\quad \gamma_5S(t_{10}-t_2)\gamma_5\Delta(t_2)=A,$$

and let the expression in double brackets, $\{\{\ \ \}\}$, in Eq. (10) be equal to A'. Then among others the following are divergent terms separated:

$$\delta\lambda\int dt_2\Delta(t_2)+Af^2\int dt_1$$
$$\times\mathrm{Trace}\,[\gamma_5S(t_1-p)\gamma_5S(t_1)S(t_1)\gamma_5]+A'. \quad (11)$$

The integral with Af^2 as coefficient will be canceled by being combined with the integral arising from the nucleon mass renormalization term $-\delta\kappa_0\bar{\psi}\psi$ in the hamiltonian, while A' is similarly canceled by a proper choice of $\delta\kappa^2$ in the term $-\delta\kappa^2\phi^2$. To cancel $\delta\lambda\int dt_2\Delta(t_2)$ we can combine it with a meson self-energy graph obtained by replacing the M part by a 4-vertex as we have so far been doing; but now we notice that

$$A'=A''-\delta\lambda\int dt_2\Delta(t_2),$$

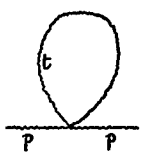


FIG. 6.

where A'' is that value of the meson mass renormalization constant which it would have if M parts were not divergent in the theory; i.e.

$$A''=f^4\int dt_1dt_2\,\mathrm{Trace}\,[\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\gamma_5S(t_1-t_2)\gamma_5S(t_1)\Delta(t_2)-\gamma_5S(t_1-p_0)\gamma_5S(t_1)$$
$$\times\{\gamma_5S(t_{10}-t_2)\gamma_5+\langle(t_1-t_{10})\partial/\partial t_1\rangle_{t_1=t_{10}}$$
$$\times(\gamma_5S(t_1-t_2)\gamma_5)\}S(t_1)\Delta(t_2)],$$

so that (11) can be written as

$$A''+Af^2\int dt_1\,\mathrm{Trace}\,[\gamma_5S(t_1-p)\gamma_5S(t_1)S(t_1)].$$

Further, since $\delta\lambda$ is not a function of p, the divergent term separated corresponding to t_1 gives no contribution to the charge renormalization constant set out in the last integral in Eq. (10) as part of the coefficient of $(p-p_0)_\mu(p-p_0)_\nu$. The result is that the divergent terms to be separated in order to obtain an absolutely convergent integral from a meson self-energy graph are precisely the same as if M-parts were not divergent in the theory, this result holding for only those M parts which are obtained by opening a single meson line in a meson self-energy graph. A further consequence of this is that the self-energy graph in Fig. 6, unlike other graphs containing 4-vertices, has to be considered separately and is not like others a compensatory graph.

The reason why the result shown to hold true for the particular example can be generalized is this. Let us first consider a meson self-energy graph belonging to any category $[n]$. The meson and nucleon lines or the

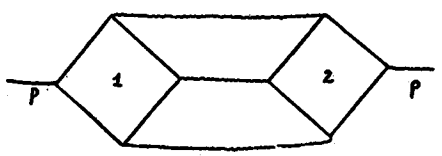


FIG. 7.

vertices in $[n]$ do not contain any S or V insertions so far. In order to remove the divergence from the M part formed by opening a meson line *we must associate one*

[9] The divergent terms shown here corresponding to t_1t_2 integration are of the form $A'+B(p^2+k^2)+C[p_0(p-p_0)]^2$ where C can be shown to be a finite constant. $C[p_0(p-p_0)]^2$ is thus properly to be included in the convergent part of Π^*.

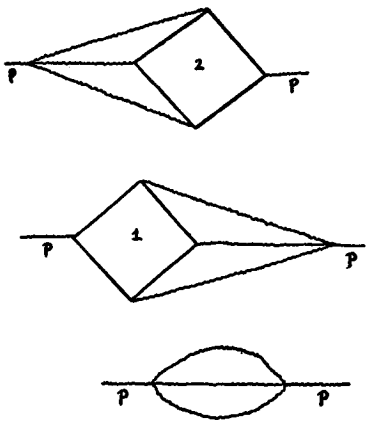


FIG. 8.

of the basic variables, say, t_1, with the particular meson line being opened. The true divergence corresponding to the subintegration $t_2t_3\cdots t_k$ is a constant ($\delta\lambda$) and the reduced integral is necessarily $\int\Delta_F(t_1)d^4t_1$. Thus, when the final true divergence over $t_1t_2\cdots t_k$ is separated, $\Delta_F(t_1)$ not containing the external momentum p, contributes $-\delta\lambda\int dt_1\Delta_F(t_1)$ to A' and makes no contribution to the charge-renormalization constant at all, precisely as in the particular example. The effect is the same as if M parts were not divergent in the theory, and the mass renormalization constant were not A' but A''.

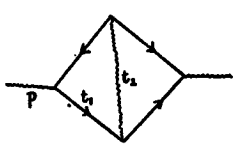


FIG. 9.

It must be emphasized that the above result holds only for those M parts obtained by opening one meson line. We shall call such M parts "final" M parts. Thus, for example, the self-energy graph shown in Fig. 7 containing the M parts with the loops marked 1 and 2 is necessarily to be combined with three other graphs in Fig. 8 just as in Sec. V(A).

So far we have restricted our considerations to graphs belonging to a category $[n]$ in which S and V parts have not been inserted. To bring out the new features involved we consider another example. Figure 9 shows a meson self-energy graph belonging to $[2]$ in which the M divergence is associated with the t_1 loop. In the line t_2 we insert the self-energy graph in Fig. 5, which is itself a modification of the graph in $[1]$. The resulting graph is shown in Fig. 10. M divergences are now

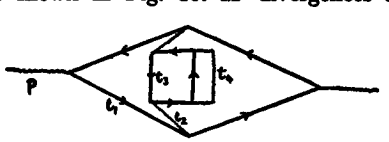


FIG. 10.

associated with the subintegrations over t_3, t_1, t_1t_3, and $t_1t_2t_3$. Considering the line t_2 by itself, from the arguments given before it is obvious that we need not make an explicit separation of the M divergence corresponding to t_3. If now the "mass-renormalization constant" corresponding to the line p is considered, it will be seen that not only is it unnecessary to separate the divergence corresponding to $t_1t_2t_3$ subintegration (for which only one variable t_4 is held fixed), but the same applies for the t_1 and t_1t_3 subintegrations. The reasoning for all cases is similar. For example, the true divergence over t_1 is obtained by putting, among other variables, p equal to p_0 *wherever* it occurs in the integrand. When the mass renormalization constant corresponding to line p is being separated, its part corresponding to t_1 is the same as above, no further change being necessary, since p has already been replaced by p_0. Thus, $A'=A''-$true divergence over $t_1\times$the reduced integral over $t_2t_3t_4$ $-$true divergence over $t_1t_3\times$reduced integral over t_2t_4 $-$true divergence over $t_1t_2t_3\times$reduced integral over t_4. Similarly, we can see that the *charge* renormalization constant is precisely the same as if these M parts were not divergent in the theory. From our definition, that M corresponding to the subintegration t_3 was the "final" M part contained in the meson self-energy graph (of Fig. 5) inserted in line t_2, while t_1 was the "final" M part for the line p in Fig. 9. Because of the insertion in t_2, new M divergences t_1t_3, $t_1t_2t_3$ arise; since they can be associated with p, we shall call the divergences corresponding to t_1, t_1t_3, and $t_1t_2t_3$ also "final" divergences associated with the line p in Fig. 10. This is an extension of the definition of "final" M-divergence.[10] In general, the result can be stated thus: given a meson self-energy graph of category $[n]$, the insertion of a meson self-energy graph with a number of final M divergences in any of its meson lines will increase the number (and complexity) of the M divergences in the entire graph. It will always be possible to associate these new divergences as the final M-divergences corresponding to some meson line, making it possible to prove that no explicit separation of corresponding M divergences is needed. We have the result that the mass and charge renormalization constants for meson self-energy graphs are precisely the same as if these "final" M parts were not divergent.

The same results (only simpler to prove) hold for insertions of S and V parts in the nucleon lines and vertices of graphs of categories $[n]$.

The only meson self-energy graph which still needs to be considered is the graph in Fig. 6. All other meson self-energy graphs with 4-vertices act as compensatory graphs like those in Fig. 8. The corresponding integral is

$$-\delta\lambda(f_1)\int\Delta_F'(t)dt=-M_d(f_1)Z_3^{-1}\int\Delta_{F1}'(t)dt,$$

[10] Analytically a final M-divergence inside a meson self-energy graph may be defined such that its reduced integral is independent of the external momentum of the meson line.

so that finally we choose

$$\delta\kappa^2 = -(1/\pi i)Z_2^{-1}(f_1)\left[A''(f_1) - M_d(f_1)\int \Delta_{F1}'(t, f_1)dt\right].$$

This completes the proof.

On account of the results above regarding "final" M-divergences, it will be found expedient to associate basic momenta with nucleon rather with meson lines in graphs belonging to any category $[n]$. This choice will remove all the significant divergences with a single choice of basic variables.

The case of the meson self-energy has been treated at length because it will serve as a prototype for renormalization in other theories such as scalar electrodynamics, where both C and M parts are divergent and it is not possible to make an unambiguous insertion of S parts in meson or photon lines.

The general rule for the separation of divergences given in Sec. III will also apply to other theories such as scalar electrodynamics; but the considerations given in the remarks will not, because both C and M parts are divergent in that theory, leading to overlaps of considerably greater complexity. It is, in fact, in scalar electrodynamics that the power of the general rule formulated in Sec. III exhibits itself. General considerations on that problem will be published shortly.

The author is deeply indebted to Mr. F. J. Dyson for indicating the considerations of Sec. II in an extremely helpful discussion, without which this work would not have been possible, and to Dr. N. Kemmer for continual help and encouragement. The tenure of a scholarship from the Education Department of the Government of Punjab, Pakistan is gratefully acknowledged.

Reprinted from REVIEWS OF MODERN PHYSICS, Vol. 23, No. 4, pp. 311–314, October, 1951

The Renormalization of Meson Theories*

P. T. MATTHEWS† AND ABDUS SALAM‡

Institute for Advanced Study, Princeton, New Jersey

A brief account is given of Dyson's proof of the finiteness after renormalization of the matrix elements for scattering processes (*S*-matrix elements) in electrodynamics (interaction of photons and electrons). It is shown to which meson interactions this proof can be extended and some of the difficulties which arose in this extension are discussed.

THE recent developments in quantum electrodynamics (the interaction of photons and the electron-positron field) associated with the names of Tomonaga, Schwinger, and Feynman culminated, as far as the theory of the renormalization of mass and charge is concerned, in the work of Dyson[1] published in 1949. Combining Feynman's technique[2] of depicting field events graphically and Schwinger's invariant procedure of subtracting divergences,[3] Dyson proved two very important results. He showed first that if calculations are made to any arbitrarily high order in the charge in a perturbation expansion, three and only three types of integrals can diverge; and, secondly, that a renormalization of mass and charge would suffice completely to absorb these divergences. This theory has proved to be in very close agreement with experiment.[4]

An obvious problem after Dyson's program was completed was to extend his considerations to the various meson theories, and to see if analogous results could be derived for any of them. This work has now been finished and it is the purpose of this note to demonstrate some of the difficulties which arose and to summarize the main results. It should be emphasized that we are concerned here with the purely mathematical problem of seeing which meson theories can be made self-consistent in this way. Very little will be said about the relation of such a theory to experiment.

Before going on to consider meson theories in detail, however, we briefly recall Dyson's procedure. It is now well known[1,2] that the matrix element for any given scattering process (*S*-matrix element) in electrodynamics can be written down directly as an integral in momentum space by drawing a graph, the integrand being obtained by writing the propagation functions $S_F(p)$ and $D_F(p)$ for the electron and the photon lines[5] and the factor $e\gamma_\mu$ (charge times a Dirac matrix) for the vertices of the graph. By considering the integrals thus obtained, Dyson showed that the over-all[6] degree of divergence of a particular graph could be estimated simply by counting its external lines. Let E_f denote the number of external fermion (we use the term fermion for any spin half particle) and E_p the number of external photon lines. The integral corresponding to a graph can diverge only if

$$\tfrac{3}{2}E_f+E_p<5. \qquad (1)$$

This basic inequality shows that there are only a finite number of *types* of graph that can introduce divergences in the theory. These are the electron and photon self-energy graphs and vertex parts, simple examples of which are given in Fig. 1 (a, b, and c). Another possible type of divergent graph is the scattering of light by light (Fig. 1d), but this proves to be convergent owing to the gauge invariance of the theory.[7] There are also potentially divergent graphs with one or three external photon lines but these can be excluded by an argument based on charge symmetry. The graphs a, b, and c are thus typical of the only types of divergence in the theory and it is clear that if these divergences can be removed

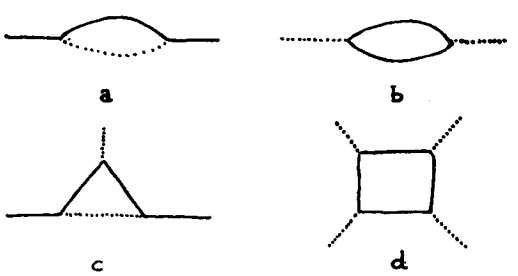


FIG. 1. Dotted lines are photons and full lines are electrons.

* The following note was read as an invited paper at the Schenectady Meeting of the American Physical Society, June 16, 1951.

† Now at Clare College, Cambridge, England.

‡ Now at Government College, Lahore, Pakistan.

[1] F. J. Dyson, Phys. Rev. 75, 486, 1736 (1949).

[2] R. P. Feynman, Phys. Rev. 76, 749, 769 (1949).

[3] J. Schwinger, Phys. Rev. 74, 1439 (1948); 75, 651 (1949).

[4] Notably it explains the Lamb shift (see W. E. Lamb and R. C. Retherford, Phys. Rev. 79, 549 (1950) for references to published papers; more accurate calculations are in progress) and the anomalous magnetic moment of the electron. P. Kush and H. A. Foley, Phys. Rev. 74, 750 (1948). J. Schwinger, Phys. Rev. 73, 415 (1948). R. Karplus and N. M. Kroll, Phys. Rev. 77, 536 (1950). Koenig, Prodell, and Kusch, Phys. Rev. 687 (1951).

[5] These are the Green's functions which express the (casually correct) influence of the fields at different points upon each other. M. Fierz, Helv. Phys. Acta 23, 731 (1950). ($D_F=D_c$ in Fierz's paper. $D_F=\delta_+$, $S_F=K_+$ in Feynman's notation.)

[6] By "over-all" degree of divergence is meant the degree of divergence of the integral over all variables, for a graph for which the integration over any lesser number of variables is finite (or has been made finite by suitable subtractions). The integrals are complex but the degree of divergence can be determined by counting powers in the integrand. For electrodynamics this can be expressed in terms of the number of external lines only.

[7] J. C. Ward, Phys. Rev. 77, 293 (1950).

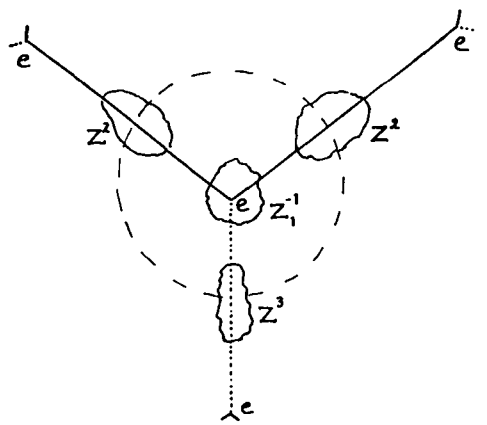


FIG. 2. The circle encloses the "basic unit" out of which all (nonoverlapping) graphs are built.

by mass and charge renormalization the entire S-matrix can be made convergent.

The graphical technique can easily be extended to the meson-nucleon interactions and a similar analysis can be made of the possible types of divergence.[8] These interactions fall into two distinct classes. For the (pseudo-) scalar interaction of (pseudo-) scalar mesons with nucleons the necessary condition for divergence is

$$\tfrac{3}{2}E_f+E_m<5, \tag{2}$$

where E_m is the number of external meson lines. For all other theories, and this includes meson pair theories and the Fermi type coupling of four fermions, the condition is of the form,

$$E<5+n, \tag{3}$$

where E is a linear function of E_f and E_m, and n is the number of vertices in the graph. For such cases there are, in general, an infinite number of types of divergence (since a divergent graph for any process can be found by making n large enough), and there is no hope of removing them by the renormalization of a finite number of constants. One exception to this rule is the *vector coupling* of *neutral vector* mesons. This has been shown to be equivalent to electrodynamics as far as divergences are concerned and is therefore renormalizable.[9] It is not, however, of much physical interest.

Turning now to the theories of spin zero particles satisfying (2), it is clear that the possible types of divergence are the same as in electrodynamics. For charged mesons the parts with one or three external meson lines are excluded by charge conservation—since the charge entering any part must leave again; for neutral pseudo-scalar mesons it can be shown that such parts vanish identically by considering the space reflection properties of the field. For neutral scalar theories these parts give rise to genuine divergences. All these theories also have genuine divergences from the scattering of mesons by mesons, Fig. 1d, in addition to the self-energy and vertex parts of electrodynamics. For this class renormalization has, on the face of it, a good chance of success and we will consider it in some detail.

Dyson based his general proof for electrodynamics on the important observation that the three divergent types of graph, when appearing as parts of a larger graph, can each be treated as modifications of single lines or vertices. One can thus define functions, analogous to the propagation functions, which include the effects of all possible self-energy parts in a line, or vertex parts at a vertex. These functions are called

$$S_F',\ D_F' \text{ and } \Gamma_\mu.$$

They are, of course, divergent. The finite, physically significant parts of these functions, S_{F1}', D_{F1}', and $\Gamma_{\mu 1}$, are obtained by subtracting infinite constants[10] in a covariant way. Dyson went on to show that after a suitable mass renormalization, these finite and corresponding infinite functions are related as follows:

$$\begin{aligned} S_F'&=Z_2S_{F1}',\\ D_F'&=Z_3D_{F1}',\\ \Gamma_\mu&=Z_1^{-1}\Gamma_{\mu 1}, \end{aligned} \tag{4}$$

where the Z's are infinite constants. These remarkable equations thus reinterpret the *subtraction* of infinite constants as the extraction of infinite constant *multipliers*.

Now consider any process, not itself one of the three divergent types. To calculate the matrix element for this process we can draw just those graphs corresponding to it which contain no self-energy or vertex parts in their lines and vertices—the "irreducible" graphs. The integral corresponding to such a graph is convergent by (1). (It satisfies the condition of reference 6 since, by definition, it contains no divergent part.) The integrand is a function of S_F D_F and $e\gamma_\mu$, say, $e^nI(S_F, D_F, \gamma_\mu)$, where n is again the number of vertices in the graph. All the radiative corrections to this graph are obtained unambiguously by making all possible self-energy and vertex part insertions in its lines and vertices. This corresponds analytically to replacing S_F by S_F', D_F by D_F', and γ_μ by Γ_μ in I. The integral now diverges, but using Dyson's equations (4) we have

$$e^nI(S_F', D_F', \Gamma_\mu)=e^nF(Z_1, Z_2, Z_3)I(S_{F1}', D_{F1}', \Gamma_{\mu 1}) \tag{5}$$

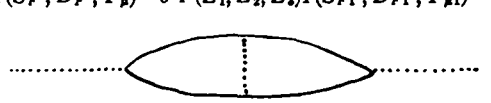


FIG. 3.

[8] P. T. Matthews, Phil. Mag. 41, 185 (1950). This paper contains errors. See P. T. Matthews, Phys. Rev. 80, 293 (1950), first paragraph.

[9] P. T. Matthews, Phys. Rev. 76, 1254 (1949). There is also a renormalizable mixture proposed by Feynman of pseudo-scalar and pseudo-vector mesons with the same mass and coupling constant and pseudo-vector coupling. See D. B. Beard and H. A. Bethe, Phys. Rev. 83, 1106 (1951).

[10] An infinite constant is an expression such as $\int_0^\infty x^{-1}dx$. Such constants are equated when the integrands are identical.

for the whole contribution from the graph, including all corrections. The form of the factor $F(Z)$ can be seen by considering a typical vertex of any irreducible graph. The self-energy and vertex parts can be inserted and give rise to the Z factors, as indicated in Fig. 2. The factors *associated with each* e are thus $Z_1^{-1}Z_2Z_3^{\frac{1}{2}}$. (One takes the root of the line factors because each is shared by the charge factors at each end of the line. External lines require special consideration.) Thus,

$$e^nF(Z)=(eZ_1^{-1}Z_2Z_3^{\frac{1}{2}})^n=e_1^n, \tag{6}$$

where e_1 is defined by the final equality. If we now equate e_1 to the experimental value of the charge, the right-hand side of (5) becomes

$$e_1^nI(S_{F1}', D_{F1}', \Gamma_{\mu 1}),$$

and we thus obtain from each irreducible graph a finite constant times an absolutely convergent integral as its contribution to the required matrix element. The physically significant contributions of all possible graphs have been included.

This is a beautiful and ingenious idea. The whole difficulty of the proof lies in establishing the relations (4), which must account correctly for the corrections to the corrections; that is, graphically, for the self-energy and vertex parts in the lines and vertices of such parts themselves. One can proceed again by modifying the lines and vertices of irreducible self-energy and vertex graphs, but the prescription is no longer unambiguous. Consider, for example, the graph of Fig. 3. This can be regarded as a modification of Fig. 1b by the insertion of a vertex part at either of the two vertices. The two vertex parts overlap on the center line. An extension of the subtraction procedure is required to derive the convergent, physically significant remainder, and the occurrence of such "overlaps" considerably complicates the derivation of equations (4). Fortunately, this is the only type of overlap which occurs in electrodynamics and Dyson was able to deal with it by special methods.[11] However, this proves to be the central problem in the extension of Dyson's methods to meson theories and it has to be treated in its most general form.[12]

Returning to the charged scalar and the charged or neutral pseudo-scalar theories satisfying (2), we have just one extra type of divergence to deal with (Fig. 1d). Since this is not a modification of any vertex, we produce such a vertex by adding a contact interaction between mesons, $(\lambda+\delta\lambda)\phi^4$, to the lagrangian. Here the idea is that $\delta\lambda$ should be an infinite constant chosen to cancel the new type of divergence, and λ is finite, to be fitted to experiment. In order to prove that this can be done consistently, it must be shown that $\delta\lambda$ can be made to cancel the divergences from the new graphs which the new vertex itself introduces. It must be shown further that with the inclusion of this term, it is still possible to define the Z factors, in spite of more complicated overlaps. This is the harder problem. The kind of difficulty which appears can be seen from the comparatively simple graph, Fig. 4, where meson-meson scatterings occur overlapping with vertex parts inside a self-energy graph.

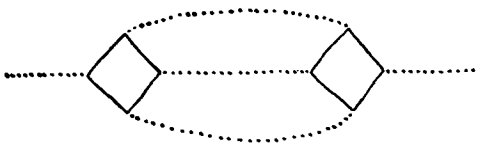


FIG. 4.

All this can in fact be done[12] but before going any further we will consider the meson-photon interactions.[13] Here a very similar situation is found with one extra complication. For spin one mesons the condition for divergence is again of the n dependent form (3) and clearly there is no hope. For spin zero mesons we have

$$E_m+E_p<5, \tag{7}$$

which is slightly less restrictive than condition (2). We end up with the same types of genuine divergence as in electrodynamics (where dotted lines are now photons, and full lines, mesons) with the addition of meson by meson scattering ($E_m=4$), and parts corresponding to the Compton effect ($E_m=2$, $E_p=2$). The hamiltonian, however, besides terms linear in e, contains a term $e^2A_\mu^2\phi^2$, giving rise to vertices of just this latter type, and the new Compton divergences can be regarded as modifications of them. We can hope to define a new factor Z_4^{-1} associated with them. We again have to introduce the extra term $(\lambda+\delta\lambda)\phi^4$ with its attendant difficulties. After deriving the Z's, we must show further that some combination of Z factors appears consistently with the e^2 vertices and that this is consistent with the Z factors at the e vertices. This requires that certain relations exist between the Z factors.[14] The overlap problem is now even more complicated. For example, the graph of Fig. 4 now contains twelve overlapping divergences. In spite of this, this program can be carried through and the whole theory rendered finite by renormalization.[15]

We do not intend here to go into any of the details of the proofs of the possibility of renormalizing these interactions, but it should be abundantly clear that the main difficulty in extending Dyson's proof to meson interactions lay in defining what infinite subtractions

[11] See also J. C. Ward, Proc. Phys. Soc. (London) 64, 54 (1951).

[12] Abdus Salam, Phys. Rev. 82, 217 (1951).

[13] P. T. Matthews, Phys. Rev. 80, 292 (1950); F. Rohrlich, Phys. Rev. 80, 666 (1950). (Rohrlich's paper contains an incorrect subtraction procedure for overlapping divergences in terms of an arbitrary "hierarchy.")

[14] *Assuming* the relations of form (4), it is easy to derive the "basic unit" for an e^2 vertex (as in Fig. 2) and to show that it will give rise to a factor $e^2Z_4^{-1}Z_2Z_3$. For consistent renormalization, we must show $Z_4^{-1}Z_2Z_3=(Z_1^{-1}Z_2Z_3^{\frac{1}{2}})^2$. In fact, $Z_1=Z_2=Z_4$ and the only genuine charge renormalization comes from Z_3 (photon self-energy parts). This is also true of electrodynamics where $Z_1=Z_2$; J. C. Ward, Phys. Rev. 78, 182 (1950).

[15] Abdus Salam (to be published in Phys. Rev.).

must be made from overlapping divergent parts so as to leave a unique, covariant, and absolutely convergent remainder. Only after this has been done can one begin on the further problem of interpreting these subtractions as infinite Z-type multiples and deriving relations between the Z factors so defined.

We content ourselves with indicating the correct subtraction procedure.[12] Consider an integral over, say, three basic variables with integrand,

$$I(t_1, t_2, t_3).$$

One must first consider the integration over t_1 alone. If this diverges one must subtract the divergent part of this integrand, as defined by Dyson, multiplied by the rest of the integrand left unchanged. That is,

$$I(t_1, t_2, t_3)-D(t_1)R(t_2, t_3).$$

This must be done *independently* for t_2 and t_3 giving

$$I(t_1, t_2, t_3)-D(t_1)R(t_2, t_3) -D(t_2)R(t_3, t_1)-D(t_3)R(t_1, t_2). \quad (8)$$

Now one considers the integral over pairs of variables. For the t_1t_2 subintegration the integrand to be considered consists of the first three foregoing terms, for the t_2t_3 integration the first, third, and fourth terms, and so on, the remaining term in each case being left unchanged. The divergent terms corresponding to these integrations, which have to be subtracted, are denoted by

$$\sum_{1,2,3} D(t_1, t_2)R(t_3). \quad (9)$$

The final subtraction, $D(t_1t_2t_3)$, is obtained by applying Dyson's procedure to the whole expression (8)–(9). This leaves behind a remainder which is convergent over all possible subintegrations and is the physically significant contribution of the graph.

A general proof has been found that, for renormalizable theories, this prescription yields a finite remainder of the required form,[16] and, for such interactions, can be interpreted in terms of Z-factors.[17] The procedure looks complicated but the idea is essentially simple. The difficulty, as in all this work, is to find a notation which is both concise and intelligible to at least two people of whom one may be the author.

[16] Abdus Salam, Phys. Rev. 84, 426 (1951).

[17] A general discussion of the extraction of Z factors is to be found in the introduction to reference 15.

We can summarize this brief survey by saying that the simplest renormalizable theory which allows for both charged and neutral mesons is the *pseudo-scalar* in *pseudo-scalar coupling* with nucleons. This theory requires just the additional interaction $(\lambda+\delta\lambda)\phi^4$. It has been found that one can consider the combined interaction of the *meson, nucleon, photon, and electron-positron* fields without introducing any essentially new complications.[18] The only other such renormalizable theory is that of *scalar* mesons with *scalar interaction*. This requires another additional term $(\chi+\delta\chi)\phi^3$ to cancel divergences from parts with three external meson lines.

It is encouraging, at least, that these purely theoretical arguments lead to the conclusion that the meson in direct interaction with the nucleon is of spin zero, in agreement with experiment. Beyond this very little can be got from the theory in its present form. Calculations for scalar theory lead to a small coupling constant but definite disagreement with experiment. Pseudo-scalar theory, on the other hand, requires so large a coupling constant to give the correct orders of magnitude that the perturbation expansion is invalidated and no comparison with experiment is possible. The problem of how to calculate cross sections by methods independent of the size of the coupling constant is a formidable one, but it seems that it must be solved before any real advance can be made in the field theoretic interpretation of nuclear events.

For theories which are not renormalizable finite cross sections have so far only been obtained by arbitrary subtraction procedures which are either not self-consistent, or incompatible with renormalization and thus in conflict with experiment.[19]

[18] P. T. Matthews, Phys. Rev. 80, 292 (1950); Phil. Mag. 42, 221 (1951); Abdus Salam, Phys. Rev. 79, 910 (1950).

[19] For example Heitler's theory of radiation damping [W. Heitler, Proc. Cam. Phil. Soc. 37, 291 (1941) and numerous subsequent papers] which gives no Lamb shift and no anomalous nucleon magnetic moments; or "regularization" discussed recently by one of us [P. T. Matthews, Phys. Rev. 81, 936 (1951)]. An interesting subtraction procedure has recently been suggested by Ning Hu [Phys. Rev. 80, 1109 (1950)], but owing to the appearance of the coupling constant in the denominator of Ning Hu's propagation function, there seems no possibility of interpreting these subtractions as renormalizations. Further, Ning Hu's modified propagation function is very similar (but more general—it is not an integral function) to those discussed and discarded by A. Pais and G. E. Uhlenbeck [Phys. Rev. 79, 145 (1950)], and it seems probable that it is open to the same objections.

Reprinted from THE PHYSICAL REVIEW, Vol. 84, No. 3, 426–431, November 1, 1951

Divergent Integrals in Renormalizable Field Theories

ABDUS SALAM*
Institute for Advanced Study, Princeton, New Jersey
(Received June 22, 1951)

The proofs of the renormalization of the theories of scalar and pseudoscalar mesons in scalar interaction with nucleons and with the electromagnetic field, given previously by the author, depended on a certain subtraction procedure. A general proof is here given that this subtraction procedure, when applied to a divergent integral in the S matrix of a renormalizable field theory, does, in fact, lead to an absolutely convergent, covariant, and unique remainder.

I

SINCE Dyson's[1] proof of the possibility of absorbing all divergences in spinor electrodynamics, by renormalizing mass and charge, it has been possible to extend his considerations and to renormalize the theories of scalar and pseudoscalar mesons in (scalar) interactions with the nucleons, and with the electromagnetic field.[2] The proofs in these cases depend upon the success of a generalized procedure[3] for subtracting divergences from the integrals occurring in these theories. This procedure was outlined in 1, Sec. III, and it was asserted that after the subtractions are carried out in a prescribed manner, the remainder is an absolutely convergent integral in a strict mathematical sense. A proof of this statement is given in this note. The conditions under which such a procedure succeeds brings out clearly the difference between theories which can be renormalized and those for which the concept of renormalization of constants has proved inadequate.

Regarding the general structure of the graphs and the corresponding integrals occurring in the matrix elements of Dyson's S matrix, we note the following:

(i) In a connected Feynman-Dyson graph consisting of N vertices and F internal lines, $n=F-N+1$ lines can be chosen such that the momenta corresponding to the remaining $(N-1)$ lines can be expressed as linear sums of the momentum variables $l_1, l_2, \cdots l_n$ which can be assigned to these n lines. Thus, the graph gives rise to an n-fold integral $I(n)$ over these basic variables.[4]

(ii) For all "renormalizable" theories (i.e., the theories for which the number of primitive divergent graphs is governed by conditions of the type $E<a$, E being a linear sum of the number of external lines and a being a fixed constant), not only is the number of primitive divergent graphs limited, but further, perhaps more important, the degree of divergence of the integrals[5] corresponding to these graphs does not depend on the order of the graphs.

(iii) We are interested in the absolute and not the conditional convergence of our final expressions; that is to say, their convergence must be independent of the choice of basic variables. Irrespective of the choice of basic variables, all the possible divergent subintegrations can be detected graphically by seeing which parts of the graph satisfy the condition $E<a$, where E is the number of external lines of the part in question. By a "wrong" choice of basic variables, it is possible to make some of the integrals corresponding to these parts conditionally convergent; for example, a subintegration[6] which conditionally converges for l_1 and l_2 integrations may diverge if $l_1'=l_1+l_2$ is chosen as a basic variable. The proof the effectiveness of the subtraction procedure given in Sec. II depends on the following lemma

Lemma 1:—There exists at least one "correct" choice of basic variables, such that every possible genuinely divergent part of the graph has associated with it a divergent subintegration over one or more of the basic variables.

This lemma is proved in Sec. III. In any graph there may be more than one set of "correct" variables. It is also shown in Sec. III that the finite remainder after the proposed subtractions is the same for any "correct" set. It will be seen in Sec. III that the lemma is true under extremely restrictive conditions. These happen to be precisely the conditions which renormalizable field theories satisfy.

II

The detailed prescription for subtracting divergences from overlapping graphs has been given in 1, Sec. III. We write the given integral $I(n)$ as

$$\begin{aligned} I(n)=&D(l_1)R(l_2l_3\cdots l_n)+D(l_2)R(l_1l_3\cdots l_n)+\cdots \\ &+D(l_1l_2)R(l_3l_4\cdots l_n)+\cdots \\ &+D(l_1l_2l_3)R(l_4l_5\cdots l_n)+\cdots+\cdots \\ &+D(l_1l_2\cdots l_n)+I_c(l_1l_2\cdots l_n). \quad (1)\end{aligned}$$

* Now at Government College, Lahore, Pakistan.

[1] F. J. Dyson, Phys. Rev. 75, 1736 (1949).

[2] Abdus Salam, Phys. Rev. 82, 217 (1951), referred to as 1; and "Renormalized S matrix for scalar electrodynamics," to be published.

[3] This procedure was suggested by F. J. Dyson, to whom the author is deeply indebted for kind help and discussion.

[4] The domain of integration is in fact a $4n$-dimensional space; but the treatment being wholly relativistic, the 4 components of any vector are never separately considered. We, therefore, speak of n-fold rather than $4n$-fold integrations.

[5] The degree of divergence of an integral is estimated by employing the basic considerations given in Sec. V of Dyson's paper and merely counting the powers of relevant l in the numerator and denominator of the integrand.

[6] If a set of variables $l_al_b\cdots l_m$ is held fixed, the integration over the remaining variables is called a subintegration.

Here $D(t_i t_j \cdots)$ is the true divergence from the part of the graph corresponding to the $t_i t_j \cdots$ subintegration and R's represent the reduced integrals. It is to be shown that I_c, the remainder left after all terms $D \times R$ have been subtracted from $I(n)$ is absolutely convergent.

In general, the integrand of $I(n)$ is a product of "simple" factors F, H, J, etc., each "simple" factor F being a function of one single variable t_i and external momenta (p), and of overlap factors G, K, etc., which are functions of more than one variable.

In any term like $D(t_1)R(t_2 t_3 \cdots t_n)$, $D(t_1)$ is a divergent constant; analytically, $D(t_1)$ is a divergent constant obtained by an operation $T(t_1)$ on the t_1 subintegration, the integrand of this subintegration consisting of all those factors F, G, etc., in the integrand of $I(n)$ which contain t_1. All momenta other than t_1 in these factors are treated as "external." There is one operator $T(t_1)$ for each type of divergence in the theory; this operator may correspond, for example, to giving all the "external" momenta special "free-particle" values or may also involve differentiations with respect to "external" momenta before such "free" values are substituted if the divergence over t_1 subintegration is more than logarithmic. In practice, the operation T may be even more complicated, but in all cases (so far developed) T acts by changing the values of "external" momenta. In actual fact, the result of operation $T(t_1)$ is not merely $D(t_1)$ but $D(t_1)\gamma_\mu$, $D(t_1)\delta_{\mu\nu}$, or $D(t_1)\Delta_F^{-1}$, etc., where γ_μ, $\delta_{\mu\nu}$, etc., will be called "vertex factors." These "vertex factors" are absorbed in the reduced integrals. $R(t_2 t_3 \cdots t_n)$, for example, is an integral with the integrand consisting of all factors G, H, etc., in the original integrand which do not contain t_1, multiplied by the "vertex factor" from the t_1 integration. $R(t_2 t_3 \cdots t_n)$ corresponds to a graph, called the reduced graph, which has exactly the same number of external lines as I, but is of lower order in e.

We notice at this stage that the operation

$$[1-T(t_1)]I(n)=[I(n)-D(t_1)R(t_2 t_3 \cdots t_n)]$$

can be thought of graphically as the insertion of the convergent part of the t_1 integration at the appropriate place in the reduced graph $R(t_2 t_3 \cdots t_n)$.

The general procedure for obtaining a true divergence, $D(t_1 t_2)$, for example, has been formulated in 1, Sec. III. Consider the integral

$$[1-T(t_1)-T(t_2)]I=[I-D(t_1)R(t_2 t_3 \cdots t_n) - D(t_2)R(t_1 t_3 \cdots t_n)]. \quad (2)$$

An operation $T(t_1 t_2)$ on this new integral gives $D(t_1 t_2)$. For this operation all momenta other than t_1, t_2 are "external" and explicitly the operation is only designed to make the double integration over $t_1 t_2$ in

$$[1-T(t_1 t_2)][1-T(t_1)-T(t_2)]I(n) \quad (3)$$

convergent. It was shown in 1, Sec. II, that (3) in fact proves to be convergent over t_1 and t_2 integrations considered singly as well. The proof may be repeated. Holding t_2 fixed, we rewrite (3) as

$$=[1-T(t_1)]I-T(t_1 t_2)[1-T(t_1)]I - [1-T(t_1 t_2)]T(t_2)I. \quad (4)$$

The first two terms obviously converge over t_1 integration.

We can rewrite the third term as

$$[1-T(t_1 t_2)]T(t_2)I=[1-T(t_1 t_2)]D(t_2)R(t_1 t_3 \cdots t_n).$$

$T(t_1 t_2)$ operates only on the $t_1 t_2$ integration, giving all variables "external" to the $t_1 t_2$ integration certain special values. Here, since $D(t_2)$ is a constant and does not contain any external momenta, $T(t_1 t_2)$ operates, in effect, only on that part of the integrand which contains t_1. The convergence of $[1-T(t_1 t_2)]D(t_2) \times R(t_1 t_3 \cdots t_n)$ over t_1 is clear dimensionally, since $D(t_2)$ being divergent, the degree of divergence of $R(t_1 t_3 \cdots t_n)$ in t_1 cannot exceed that of the $t_1 t_2$ divergence, which $[1-T(t_1 t_2)]$ is explicitly designed to remove. In the example considered in 1, Sec. II, $T(t_1)$ and $T(t_2)$ represent extraction of (logarithmic) vertex part divergences, while $T(t_1 t_2)$ is designed, in the first place, to extract (linear) self-energy divergences. $D(t_2)$ thus being a logarithmic divergence, the t_1 integration in $R(t_1 t_3 \cdots t_n)$ can at most be linearly divergent. The operation $[1-T(t_1 t_2)]$ therefore suffices to produce convergence over t_1 in $[1-T(t_1 t_2)]T(t_2)I$. This is an important point for the subsequent proof.

In the general case I_c is convergent for the over-all n-fold integration by its very construction, since it is obtained from Φ defined in Eq. (5) below by an operation $[1-T(t_1 t_2 \cdots t_n)]$, which operation subtracts from Φ the true divergence $D(t_1 t_2 \cdots t_n)$:

$$\Phi=I-D(t_1)R(t_2 \cdots t_n)\cdots-D(t_1 t_2)R(t_3 \cdots t_n)\cdots$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-D(t_1 t_2 \cdots t_{n-1})R(t_n)\cdots. \quad (5)$$

We now want to show that I_c is also convergent over all subintegrations $t_1, \cdots, t_1 t_2, \cdots, t_1 t_2 \cdots t_{n-1}, \cdots$ as well. Combining this with the discussion given in Sec. III, this will establish that I_c is absolutely convergent in a strict mathematical sense and all mathematical operations are valid for it.

By rearranging the terms in $I_c(n)$, we shall first exhibit that $I_c(n)$ is convergent over the $(t_2 t_3 \cdots t_n)$ subintegration, explicitly. The proof proceeds inductively. Suppose that $I_c(n-1)$ obtained from any $(n-1)$ fold integral $I(n-1)$ by subtracting true divergences $\times$reduced integrals converges over all its possible subintegrations as well as over the $(n-1)$ fold inte-

gration. We rewrite Eq. (5) as follows:

$$\Phi=[I-D(t_2)R(t_1t_3\cdots t_n)-D(t_3)R(t_1t_2\cdots t_n)$$
$$-\cdots-D(t_n)R(t_1\cdots t_{n-1})-D(t_2t_3)R(t_1t_4\cdots t_n)$$
$$-\cdots-D(t_2t_3\cdots t_n)R(t_1)]$$
$$-[D(t_1)R(t_2t_3\cdots t_n)-D(t_1t_2)R(t_3t_4\cdots t_n)\cdots$$
$$-D(t_1t_n)R(t_2\cdots t_{n-1})\cdots-D(t_1t_2t_3)R(t_4t_5\cdots t_n)\cdots$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-D(t_1t_3\cdots t_n)R(t_2)-D(t_1t_2\cdots t_n)R(t_3)\cdots$$
$$-D(t_1t_2\cdots t_{n-1})R(t_n)]. \quad (6)$$

Then we have

$$I_c(n)=[1-T(t_1\cdots t_n)]\Phi. \quad (7)$$

If t_1 is held fixed, it acts as an "external" momentum vector for the terms in the first bracket. The integrand in $R(t_1)$ contains (apart from the "vertex factor") those factors in the integrand of $I(n)$ which do not contain any other basic variable except t_1, and these factors are therefore common to all the terms in the first bracket in Eq. (6). Apart from these, the first bracket is precisely $I_c(n-1)$ and, consequently, convergent over $t_2t_3\cdots t_n$ integration, by assumption.

To exhibit the explicit convergence of the second bracket, we derive an alternative expression for the true divergences.

Consider $D(t_1t_2)$. We write

$$D(t_1t_2)R(t_3t_4\cdots t_n)=T(t_1t_2)[I-D(t_2)R(t_1t_3\cdots t_n)$$
$$-D(t_1)R(t_2t_3\cdots t_n)]$$
$$=T(t_1t_2)[I-D(t_2)R(t_1t_3\cdots t_n)]$$
$$-D(t_1)D_R(t_2)\times R(t_3t_4\cdots t_n). \quad (8)$$

Here $D_R(t_2)R(t_3t_4\cdots t_n)$ is the result of the operation $T(t_1t_2)$ (which, $D(t_1)$ being a constant, is effectively now an operation on t_2 integration alone) on $R(t_2t_3\cdots t_n)$. The reduced integral $R(t_3t_4\cdots t_n)$ from this operation, obtained by omitting all factors containing t_2 in the integrand of $R(t_2t_3\cdots t_n)$ is the same as that obtained by omitting all factors containing t_1t_2 from the integrand of I. As noted before $[I-D(t_2)R(t_1t_3\cdots t_n)]$ can be thought of as the graph obtained from $R(t_1t_3\cdots t_n)$ by inserting in the appropriate place the convergent remainder $G_c(t_2)$ of the t_2 integration in I. The operation $T(t_1t_2)$ now implies that from the resulting expression another divergent constant is evaluated. For our purposes, however, the essential point is that $T(t_1t_2)[I-D(t_2)R(t_1t_3\cdots t_n)]$ is convergent over t_2. To emphasize this we write $T(t_1t_2)[I-D(t_2)R(t_1t_3\cdots t_n)]$ as $(G_c(t_2)\rightarrow\int dt_1dt_2)_0R(t_3t_4\cdots t_n)$. Here, $(G_c(t_2)\rightarrow\int dt_1dt_2)_0$ is the divergent constant mentioned above and is such that in its integral representation the t_2 integration is convergent. Its multiplier $R(t_3t_4\cdots t_n)$ is an integral with the integrand consisting of those factors in the integrand of I which contain neither t_1 nor t_2 and is thus the correct reduced integral. Hence, finally we have

$$D(t_1t_2)R(t_3\cdots t_n)=[(G_c(t_2)\rightarrow\int dt_1dt_2)_0$$
$$-D(t_1)D_R(t_2)]\times R(t_3\cdots t_n). \quad (9)$$

It may be noted here that if we take two variables t_1 and t_i such that the parts corresponding to t_1 and t_i integrations do not overlap, then we have $D_R(t_i)=D(t_i)$. (The true divergence for the t_i subintegration of the reduced integral $R(t_2\cdots t_n)$ is the same as that of the original integral $I(t_1t_2\cdots t_n)$.) It was shown in Sec. 3, 1, that for such a case we have $D(t_1t_i)=-D(t_1)D(t_i)$. Hence, it follows that $(G_c(t_i)\rightarrow\int dt_1dt_i)_0=0$, which is compatible with its graphical interpretation.

We shall prove a similar result for $D(t_1t_2t_3)$. By definition we have

$$D(t_1t_2t_3)R(t_4\cdots t_n)=T(t_1t_2t_3)[I-D(t_2)R(t_1t_3\cdots t_n)$$
$$-D(t_3)R(t_1t_2t_4\cdots t_n)-D(t_2t_3)R(t_1t_4\cdots t_n)]$$
$$+T(t_1t_2t_3)\begin{bmatrix}-D(t_1)R(t_2t_3\cdots t_n)\\-D(t_1t_2)R(t_3t_4\cdots t_n)\\-D(t_1t_3)R(t_2t_4\cdots t_n)\end{bmatrix}. \quad (10)$$

The expression in the first square bracket is convergent on account of our inductive assumption for the t_2t_3 integration. Precisely as above, we can write the result of operating $T(t_1t_2t_3)$ on the first bracket as

$$(G_c(t_2t_3)\rightarrow\int dt_1dt_2dt_3)_0R(t_4\cdots t_n).$$

By a repeated use of Eq. (9), the remaining terms in Eq. (10) can be written as

$$=-T(t_1t_2t_3)[D(t_1)\{R(t_2t_3\cdots t_n)-D_R(t_2)R(t_3t_4\cdots t_n)$$
$$-D_R(t_3)R(t_2t_4\cdots t_n)\}$$
$$+(G_c(t_3)\rightarrow\int dt_1dt_3)_0R(t_2t_4\cdots t_n)$$
$$+(G_c(t_2)\rightarrow\int dt_1dt_2)_0R(t_3t_4\cdots t_n)]$$
$$=-[D(t_1)D_R(t_2t_3)+(G_c(t_2)\rightarrow\int dt_1dt_2)_0D_R(t_3)$$
$$+(G_c(t_3)\rightarrow\int dt_1dt_3)_0D_R(t_2)]\times R(t_4\cdots t_n). \quad (11)$$

So that, finally we have

$$D(t_1t_2t_3)R(t_4\cdots t_n)=[(G_c(t_2t_3)\rightarrow\int dt_1dt_2dt_3)_0$$
$$-(G_c(t_3)\rightarrow\int dt_1dt_3)_0D_R(t_2)$$
$$-(G_c(t_2)\rightarrow\int dt_1dt_2)_0D_R(t_3)$$
$$-D(t_1)D_R(t_2t_3)]R(t_4\cdots t_n). \quad (12)$$

This can be generalized immediately by an inductive proof as follows:

$$D(t_1t_2\cdots t_r)R(t_{r+1}\cdots t_n)$$
$$=[(G_c(t_2\cdots t_r)\rightarrow\int dt_1dt_2\cdots dt_r)_0$$
$$-(G_c(t_2\cdots t_r)\rightarrow\int dt_1dt_2\cdots dt_r)_0D_R(t_2)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-(G_c(t_2\cdots t_{r-1})\rightarrow\int dt_1dt_2\cdots dt_{r-1})_0D_R(t_r)$$
$$-(G_c(t_4\cdots t_r)\rightarrow\int dt_1dt_4\cdots dt_r)_0D_R(t_2t_3)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-(G_c(t_5\cdots t_r)\rightarrow\int dt_1dt_5\cdots dt_r)_0D_R(t_2t_3t_4)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-(G_c(t_r)\rightarrow\int dt_1dt_r)_0D_R(t_2t_3\cdots t_{r-1})$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$-D(t_1)D_R(t_2t_3\cdots t_r)]\times R(t_{r+1}\cdots t_n). \quad (13)$$

All that remains now is to substitute from Eq. (13) in the expression obtained after operating $[1-T(t_1t_2\cdots t_n)]$ on the second bracket of Eq. (6). We obtain

$$
\begin{aligned}
&-D(t_1)[R(t_2t_3\cdots t_n)-D_R(t_2)R(t_3\cdots t_n)-\cdots\\
&\qquad -D_R(t_2t_3)R(t_4\cdots t_n)-\cdots\\
&\qquad \cdots\cdots\cdots\cdots\\
&\qquad -D_R(t_2t_3\cdots t_n)]\\
&-[(G_c(t_2)\rightarrow\int dt_1dt_2)_0\{R(t_3\cdots t_n)-D_R(t_3)R(t_4\cdots t_n)\\
&\qquad -D_R(t_3t_4)R(t_5\cdots t_n)-\cdots-D_R(t_3\cdots t_n)\}]\\
&-[(G_c(t_2)\rightarrow\int dt_1dt_3)_0\{R(t_2t_4\cdots t_n)\\
&\qquad -D_R(t_2)R(t_4\cdots t_n)\cdots\\
&\qquad -D_R(t_2t_4)R(t_5\cdots t_n)\cdots-D_R(t_2t_4\cdots t_n)\}]\\
&\qquad \cdots\cdots\cdots\cdots\\
&-[(G_c(t_2t_3)\rightarrow\int dt_1dt_2dt_3)_0\{R(t_4\cdots t_n)\\
&\qquad -D_R(t_4)R(t_5\cdots t_n)\cdots\\
&\qquad -D_R(t_4t_5)R(t_6t_7\cdots t_n)-D_R(t_4t_5\cdots t_n)\}]\\
&\qquad \cdots\cdots\cdots\cdots\\
&-(G_c(t_2\cdots t_n)\rightarrow\int dt_1dt_2\cdots dt_n)_0. \qquad (14)
\end{aligned}
$$

By our inductive hypothesis each expression in square brackets in the above converges for the $t_2t_3\cdots t_n$ integration if t_1 is held fixed. This proves the result. Since the inductive hypothesis assumes the convergence of $I_c(n-1)$ for all its possible subintegrations, we have succeeded in establishing the convergence of $I_c(n)$ over all subintegrations t_2, t_2t_3, $\cdots$, $t_2t_3\cdots t_{n-1}\cdots$ as well.

By different rearrangements we could prove the convergence of $I_c(n)$ over any other subintegration over $(n-1)$ variables. This shows that $I_c(n)$ is absolutely convergent, provided we can give a proof of the lemma stated in Sec. I. This is done in the next section.

III

In this section we give a proof of the lemma stated towards the end of Sec. I, on which the discussion in Sec. II depends. The proof is long, and only an outline will be given. We first investigate the condition for primitive divergents for the most general mixture of fields. Consider a graph with E_b as the total number of external boson and E_f as the total number of external fermion lines. Let N_i stand for the number of such vertices in the graph as have i lines incident; thus, i may be any integer equal or greater than two. The number of boson lines at each vertex of the type i is b_i and the number of fermion lines f_i.

If F stands for internal lines, then we have the relations,

$$E_b+2F_b=\sum b_iN_i, \quad E_f+2F_f=\sum f_iN_i.$$

We restrict ourselves to theories[7] in which the integrals for matrix-elements can be written in momentum space by setting in the integrand $\Delta_F(l)$ for each boson line, $S_F(l)$ for each fermion line, and a vertex factor of the general form t^{α_i} for each vertex of the type i. The condition for convergence of a graph, following Dyson, is derived as

$$2F_b+F_f-[4(F-N+1)+\sum\alpha_iN_i]>1.$$

Since $N=\sum N_i$, this reduces to

$$\tfrac{3}{2}E_f+E_b+\sum N_i(4-\alpha_i-b_i-\tfrac{3}{2}f_i)-4>1.$$

The upper limit to the number of primitive divergent graphs is therefore given by the following condition:[8]

$$\tfrac{3}{2}E_f+E_b<5+\sum(\alpha_i+b_i+\tfrac{3}{2}f_i-4)N_i.$$

This result shows that if for a mixture of fields the number of primitive divergent graphs does not depend on the order of the graphs considered, then this number is governed by a condition of the type $E<a$ with $a=5$. Any graph in these theories, with four or more external fermion lines or five or more external boson lines must be (at least "superficially"[9]) convergent. This implies that if we open one single line of any graph with at least two external fermion lines, and if the resulting graph is a "connected" graph (i.e., the resulting graph does not split into pieces each joined to the other by a single line), it must converge superficially. The same result holds *a fortiori* if two or more lines which leave the graph connected are opened simultaneously. It is also seen to be valid for all graphs with boson lines external except for boson self-energy graphs. In this case, opening one boson line may produce a connected graph with 4 external boson lines which may diverge logarithmically. However, precisely this was the case treated in 1(B), where it was shown that such divergences affect only the mass-renormalization constant. These divergences were called "final"[10] and it was proved quite generally that one could effectively proceed as if these divergences did not exist. We can therefore state the following as a general rule: *If in any graph one or more lines are opened such that the graph still remains "connected," the resulting graph is either superficially convergent or such that its divergence is "final" and we can effectively proceed as if the divergence doesn't exist.* Using the above result we proceed to the direct proof of Lemma 1. By "exhibiting" a part of a graph we mean opening enough lines so that the momenta

[7] This excludes β-formalism of meson theory, for which the propagation function is

$$T_F(l)=\{i\beta l-K+(1/K)[(\beta l)^2-\beta^2]\}/(\beta^2+K^2).$$

[8] For a theory to have a finite number of primitive divergents, it is clear that $\sum(\alpha+b+\frac{3}{2}f-4)N$ must equal zero. This therefore excludes all vector couplings for meson-nucleon interactions ($b=1$, $f=2$, $\alpha=1$), meson-pair theories ($b=2$, $f=2$) and Fermi types of interaction between four spinor particles ($f=4$).

[9] An integral is "superficially" convergent, if by counting the powers of l in the integrand, the integral itself is found to be convergent, while any one of its subintegrations diverges.

[10] One further instance of a "final" divergence [not noted in 1(B)] is provided by vertex parts with 2 meson and 1 external photon line. If the external photon line belongs to a 3-vertex, opening one of the internal meson lines at this vertex may produce a connected M-part. This M-divergence can be treated as "final." The reduced integral in this case is a vertex part which can be shown to be a function of the external photon momentum $(p-p')$ only. The reduced integral would therefore vanish from the familiar arguments of gauge-invariance.

associated with the external lines of the part become fixed. It is clear that any part (convergent or divergent) may be exhibited by opening all its external lines. If we can show that the number of lines, some or all of which need to be opened, to exhibit every divergent part contained in a graph, is never greater than n (the number of independent momenta at our disposal), we will have proved the lemma. For we can then associate basic momenta with these lines, and thus have a "correct" choice of basic variables.

All internal lines belonging to any graph can be divided into two mutually exclusive sets. Lines in set (a) have the property that if any one of them is opened the graph splits into two or more parts each joined to the other by a single line (called a "bridge"). All lines not belonging to set (a) fall in set (b) which thus includes lines which if opened singly leave the graph connected.

The lines in set (a) can further be divided into distinct classes; If opening a line l_a^1 splits the graph into separate pieces joined by single lines (bridges) l_a^2, l_a^3, $\cdots$, it is obvious that the result of opening l_a^2 would be to split the graph into pieces joined by bridges l_a^1, l_a^3, $\cdots$. If a basic momentum variable l is assigned to l_a^1, the lines l_a^2, l_a^3, etc., acquire momenta $l+p$, where p are linear sums of momenta of external lines. All these lines will belong by definition to class R_a. We can assign all lines in set (a) uniquely to various classes R_1, R_2, $\cdots R_s$. No line (a) can belong to more than one class at a time, and each class must contain at least two lines. If those vertices of the graph at which external lines are incident are 3-vertices, each internal line at such a vertex certainly belongs to some one of these classes. For distinction we shall label such classes as R^{ext}. If R^{ext} consists of just the two internal lines incident at the external vertex defining R^{ext}, then the effective result of opening one of these external lines is one single connected graph with external lines one more than the original graph. We further distinguish such R^{ext} by writing these as $\mathbf{R}^{\text{ext}}$.

The connection of these sets and classes with divergences in a graph is obvious. On account of the rule stated above, opening a line (b) leaves the graph connected and therefore at least superficially convergent, while a line (a) may split the graph into pieces at least one of which has external lines satisfying $E<a$. To "exhibit" these divergences, it is therefore essential that a basic momentum variable be associated with this line (a). If (a) belongs to a class R_a, an assignment of the basic variable to any other line in R_a is, from the definition of a class, clearly an equivalent and a "correct" procedure. As a first step towards the proof of Lemma 1 we show that s, the number of effective classes, which can split the graph into pieces all or some of which satisfy $E<5$ is always $\leqslant n$.

For graphs consisting entirely of 4-vertices the result is trivial. For this case, we have $n=F-N+1$, while $2F+E=4N$. Each of the s classes must contain at least two lines; since the minimum number of lines in the classes, $2s$, must not exceed the total number of internal lines F, we have $2s\leqslant F=2n+\frac{1}{2}E-2$. For the case of divergent graphs with $E\leqslant 4$, this shows that $s\leqslant n$.

To prove the result for graphs with 3-vertices we establish Lemma 2.

Lemma 2:—The maximum number of classes s for a proper graph with n basic vectors and no external lines is $n-2$.

For the proof it is unnecessary to distinguish between boson and fermion lines. We can always obtain any graph with n basic vectors from a graph with $(n-1)$ basic vectors by adding a new line with its terminating vertices on two existing lines of the graph with $(n-1)$ basic vectors. In general, the addition of a new line does not add to the number of the existing classes except for two cases: (i) when the terminating vertices of the new line both belong to the same line of the original graph and (ii) when these vertices belong to two lines already belonging to a class. In both these cases the number of classes is increased at most by one. The proof of the lemma follows by induction, since the result is obviously true for $n=2$. We also notice that at the new terminating vertices there are at most two lines belonging to set (a), while the newly added line itself must belong to set (b). Generalizing from this, we have the result that at no vertex can there be more than two (a) lines incident. We shall have occasion to use this result.

We now consider the effect of adding external lines to such graphs. Each external line may produce at most one new class $\mathbf{R}^{\text{ext}}$, so that the general result is that for a graph with n independent vectors and E external lines

$$\max s\leqslant n+E-2.$$

We can now show that for $E=3$, at least one of classes $\mathbf{R}^{\text{ext}}$ either produces a convergent connected graph when one of its 2 lines is opened, or a "final" divergence. The former is true for $\mathbf{R}^{\text{ext}}$ formed by the external photon or meson line in vertex parts in all theories except scalar electrodynamics where this $\mathbf{R}^{\text{ext}}$ introduces, as noted before, merely a "final" divergence. Excluding this from the classes of interest we have $s\leqslant n$ for $E=3$. For the case $E=4$, opening one of the lines of an $\mathbf{R}^{\text{ext}}$ must produce a connected graph with five external lines, i.e., a subintegration which is superficially convergent. This is not true, however, if two of the R^{ext} coincide (i.e., when two external lines are adjacent), in which case the graph produced has just four external lines. Thus, in this case too the effective number of classes which exhibit divergent parts is governed by $s\leqslant n$.

With these preliminary results we now prove that we can choose a set of n lines which if opened singly or in pairs, etc., would exhibit all the divergences of the graph. Consider the case $s=n$. We assign s basic

momenta to the s lines, selected one each from the s classes R. The assignment of all momenta is complete and it remains to prove that we only need open some or all of these n lines to exhibit every possible divergence. We notice now that for this case the lines (b) acquire an additional property. By counting the lines and vertices we can show that at each vertex there are precisely two lines (a) and not more than one line (b). (We are here dealing with graphs with 3-vertices only), and also there are in all just $(n-1)$ lines (b). If we open one of these lines (b), the graph still remains "connected." Lines (a) defined for the original graph retain their property of splitting graphs even for the new graph; therefore, there are already two (a) lines at each vertex of the new graph. We have seen that at any vertex there cannot be more than two (a) lines. This shows that the (b) lines of the original graph remain the (b) lines of the new graph. The "connectedness" of the graph no longer depends on the lines (b); in other words, lines (b), which were defined in the first place on account of their property of leaving the graph connected if opened one at a time, acquire the further property that they still leave the graph connected if more than one of them are opened simultaneously.[11] The result of opening any set of these lines will therefore never be a divergent part. All divergent parts of the graphs are thus exhibited by opening one or more of lines (a); analytically, this assignment of basic momenta will insure that the entire number of divergences in the graph never exceeds the possible subintegrations at our disposal, and therefore the lemma is proved.

The uniqueness of $I_c(l_1 l_2 \cdots l_n)$ for a different "correct" choice of basic momenta follows immediately from the concept of classes. For the case dealt with above, the choice is correct only if the basic lines are selected one each from a class R. Since all other lines in a class acquire momenta $p+l$, if l is the basic line which belongs to R, all "correct" choices are completely equivalent, and obtainable one from the other by trivial transformations involving only linear sums of external momenta (p). This proves the result.

We will not deal in detail here with the proof of Lemma 1 for the case of $s<n$ where some at least of the $(n-1)$ fold integrations are "superficially" convergent. The proof requires formulation of rules for assignment of the remaining $(n-s)$ moments which is not difficult.

The proofs in this section, though tedious, are essential because it is important to realize that a subtraction procedure is worthless unless it can be shown that it leads to unique absolutely convergent results.

The author is indebted to Dr. P. T. Matthews and Dr. Res Jost for help in Sec. III and to Professor J. R. Oppenheimer for the very kind hospitality of the Institute for Advanced Study.

[11] This result is of course not true for $s<n$.

Reprinted from THE PHYSICAL REVIEW, Vol. 86, No. 5, 731-744, June 1, 1952

Renormalized S-Matrix for Scalar Electrodynamics

ABDUS SALAM*
Institute for Advanced Study, Princeton, New Jersey
(Received February 16, 1951)

By a consistent use of the concepts of mass and charge renormalization Dyson has demonstrated the possibility of constructing a divergence-free S-matrix for scattering problems in spinor electrodynamics, valid to all orders in the fine-structure constant. In this paper a proof is given of the possibility of renormalizing theories of charged scalar (or pseudoscalar) particles in the presence of the electromagnetic field. It is found that in addition to the renormalization of mass and charge, an infinite constant of direct interaction $\delta\lambda$ has to be introduced in a term $\delta\lambda\phi^{*2}\phi^2$ added to the Hamiltonian, in order to cancel consistently all divergences arising from the Møller scattering of one spinless particle by another. This, combined with the fact that theories of charged scalar and pseudoscalar mesons in scalar interaction with the nucleons can be renormalized with the same additional term in the Hamiltonian, seems to be of some significance.

1. INTRODUCTION

BY establishing the equivalence of Tomonaga-Schwinger formalism for spinor electrodynamics with that of Feynman, Dyson[1] has formulated an S-matrix theory for scattering problems in spinor electrodynamics. In this theory any real process taking place with specific virtual processes can be represented by a graph, the contribution to the matrix element for this process being an integral in the writing of which each line and each vertex of the aforementioned graph contributes a single factor. By analyzing the divergences in these integrals and giving precise rules for separating divergent parts from the (physically significant) convergent parts of these integrals, Dyson has further demonstrated that all divergences in the theory can be absorbed in the unobservable mass and charge renormalization of the theory.

For scalar electrodynamics it has been possible[2] to develop corresponding graphical methods and thus to analyze the possible divergences of the theory. It is the purpose of this paper to consider the adequacy of the concepts of renormalization for dealing with these divergences.

If $\phi(x)$ and $A_\mu(x)$ represent the meson and electromagnetic fields, respectively, the interaction Hamiltonian for mesons interacting with the electromagnetic field is

$$H_0(x)=\frac{ie}{\hbar c}A_\mu(x)\left[\phi^*(x)\frac{\partial\phi}{\partial x_\mu}-\frac{\partial\phi^*(x)}{\partial x_\mu}\phi(x)\right]-\left(\frac{ie}{\hbar c}\right)^2\phi^*(x)\phi(x)\delta_{\mu\nu}A_\mu(x)A_\nu(x)-\left(\frac{ie}{\hbar c}\right)^2\phi^*(x)\phi(x)[A_\mu(x)n_\mu]^2. \quad (1)$$

$A_\mu(x)$ creates and annihilates photons, while $\phi^*(x)$ and $\phi(x)$ are interpreted as creation and annihilation charge operators, so that ignoring the surface-dependent terms in the Hamiltonian, we have effectively[3]

$$\left\langle P\frac{\partial\phi(x)}{\partial x_\mu}\frac{\partial\phi^*(y)}{\partial y_\nu}\right\rangle_0=\tfrac{1}{2}\hbar c\frac{\partial^2}{\partial x_\mu\partial y_\nu}\Delta_F(x-y).$$

In Feynman graphs, the terms in the Hamiltonian linear in $A_\mu(x)$ lead to 3-vertices, with 2 meson lines and 1 photon line incident, while the term bilinear in $A_\mu(x)$ leads to 4-vertices, with 2 meson and 2 photon lines incident.

The contribution of this graph to the matrix[2] element appears as an integral in momentum space; in the integrand there appear[4]

(i) constant factors $\phi^*(p)$, $\phi(p)$, or $A_\mu(p)$, corresponding to each external line of the graph.

(ii) a factor $\delta_{\mu\nu}(2\pi)^{-3}D_F(p)$ for each internal photon line.

(iii) a factor $\hbar c(2\pi)^{-3}\Delta_F(p)$ for each internal meson line.

(iv) a factor $ie(\hbar c)^{-2}(2\pi)^4(p+p')_\mu\delta(p-p'+q)$ for each 3-vertex. The suffix μ gives the polarization of the photon line.

(v) a factor $-ie(\hbar c)^2(2\pi)^4\delta_{\mu\nu}\delta(p-p'+q-q')$ for each 4-vertex, where p is the momentum vector of the incoming meson, p' that of the meson leaving the 4-vertex, and q, q' refer to the photon lines incident at the 4-vertex.

(vi) the whole integral is multiplied by the number of different ways the operators can be paired off by interchanging the roles of photon operators $A_\mu(x)$ and $A_\nu(x)$ at the 4-vertices.

Thus, in general, it is possible to make a formal distinction between the factors $\delta_{\mu\nu}$ and $\delta_{\nu\mu}$ for a 4-vertex. If this distinction is not made, the integral obtained by following rules (i) to (v) must be multiplied by a "weight factor" 2 for each 4-vertex. Some exceptions

* Now at Government College, Lahore, Pakistan.

[1] F. J. Dyson, Phys. Rev. 75, 1736 (1949), hereafter referred to as D II.

[2] P. T. Matthews, Phys. Rev. 80, 282 (1950); F. Rohrlich, Phys. Rev. 80, 666 (1950). The author is indebted to P. T. Matthews for sending him a copy of F. Rohrlich's work prior to publication.

[3] P. T. Matthews, Phys. Rev. 76, 1657 (1949).

[4] These rules were already implicit in Feynman's work [R. P. Feynman, Phys. Rev. 76, 769 (1949)].

occur, however, when the weight factor is not 2, and these have been discussed in detail by Rohrlich.[2] For theoretical considerations, however, we shall always distinguish (if possible) between factors $\delta_{\mu\nu}$ and $\delta_{\nu\mu}$ for 4-vertices and no weight factor will be used.

The genuine primitive divergents[2] are of the same type as in spinor electrodynamics, namely, meson and photon self-energies and vertex parts (parts with 2 external meson lines and 1 external photon line), with the addition of logarithmically divergent graphs with two external photon and meson lines (C parts), and four external meson lines (M parts). The self-energy (S) and vertex (V) parts modify the factors from single lines or single 3-vertices, while the divergent C parts can be regarded as modifications of 4-vertices. Define $\sum_M{}^*(p)$ as the function arising from adding together all integrals corresponding to proper[5] meson self-energy parts, $\Pi^*(p)$ the corresponding sum for photon self-energy parts, $\Lambda_\mu(p, p')$ the function arising from adding the integrals corresponding to proper V parts, $\theta_{\mu\nu}(p, p', q)$ as the sum arising from adding integrals corresponding to proper C parts, and

$$D_F'=D_F+D_F\Pi^*D_F'$$
$$\Delta_F'=\Delta_F+\Delta_F\sum_M{}^*\Delta_F'$$
$$\Gamma_\mu(p, p')=(p+p')_\mu+\Lambda_\mu(p, p')$$
$$C_{\mu\nu}(p, p', q)=\delta_{\mu\nu}+\theta_{\mu\nu}(p, p', q).^6$$

Every graph G, other than a primitive divergent graph, has a uniquely defined "skeleton" which is the graph obtained by omitting all self-energy parts from the lines, vertex parts from the 3-vertices, and C parts from the 4-vertices. A graph which is its own skeleton will be called "irreducible," and a graph not containing M parts inside it "simple." With the foregoing definition of irreducibility, an irreducible graph may still contain divergent M parts.

For simple irreducible primitive divergent graphs, the forms of the functions $\sum_M{}^*$, Π^*, Λ_μ, and $\theta_{\mu\nu}$ are given as follows by invariance consideration:

$$\sum_M{}^*(p)=A+B(p^2+\kappa^2)+\Delta_c(p)(p^2+\kappa^2),$$
$$\Pi^*(p)=Cp^2+D_c(p)p^2,$$
$$\Lambda_\mu(p, p')=L(p+p')_\mu+\Lambda_{\mu c}(p, p'),$$
$$\theta_{\mu\nu}(p, p', q)=R\delta_{\mu\nu}+\theta_{\mu\nu c}(p, p', q),$$

where the suffix c stands for the convergent parts of the functions on the left-hand side of the equations.

A, B, C, L, and R are divergent constants, while the definitions of the convergent parts are such that

$$\Delta_c(p_0)=0$$

when

$$p_0{}^2+\kappa^2=0$$

and

$$D_c(0)=0$$

$\Lambda_{\mu c}(p, p')$ can be written in the form[7]

$$\Lambda_{\mu c}(p, p')=M(p+p')_\mu(p^2+\kappa^2+p'^2+\kappa^2)+F_\mu(p, p') \quad (2a)$$

where

$$F_\mu=\partial F_\mu/\partial p_\nu=\partial F_\mu/\partial p_\nu'=0 \quad \text{for} \quad p=p'=p_0$$

and in general we can write $\theta_{\mu\nu c}$ as

$$\theta_{\mu\nu c}(p, p', q)=N(p+p')_\mu(p+p')_\nu+F_{\mu\nu}(p, p', q) \quad (2b)$$

with

$$F_{\mu\nu}(p_0, p_0, 0)=0.$$

We notice that M is finite and, furthermore, that with the foregoing (unique) definition, $\Lambda_{\mu c}(p_0, p_0)\equiv 0$. Similarly, N in (2b) is a finite constant. The precise significance of these particular separations will become clear in Sec. III when we establish relations between $\sum_M{}^*$, Λ_μ, and $\theta_{\mu\nu}$.

It is possible to obtain a divergence free S-matrix if we can show:

(a) That all infinities associated with self-energy of a free-field meson can be canceled by bringing into the interaction Hamiltonian by means of a unitary transformation,[8] the mass renormalization term $-\delta\kappa^2\phi^*\phi$. The free meson field now propagates with the term $\kappa^2\phi^*\phi$ in the "free" Hamiltonian ($\kappa=mc/h$ where m is the observed meson mass).

(b) That all M divergences occurring anywhere in the theory can be consistently compensated by suitably choosing a constant $\delta\lambda$ in a term $\delta\lambda\phi^{*2}\phi^2$ which is added to the interaction Hamiltonian.[9] After the additions (a) and (b) the Hamiltonian is

$$H_1(x)=H_0(x)-\delta\kappa^2\phi^*\phi+\delta\lambda\phi^{*2}\phi^2. \quad (3)$$

The additional terms give rise to new graphs containing 2- and 4-vertices, with 2 and 4 meson lines incident, respectively. Graphs containing no such 2- or 4-vertices will be called "original."

Unlike spinor electrodynamics $H_0(x)$ contains two constants, e and e'^2, with $e=e'$. The graphs arising entirely from e-vertices (3-vertices) contain S divergences and V divergences; those from e^2 vertices (4-vertices) contain S divergences and C divergences. It would be possible to absorb these divergences in the renormalization of charge if we could show (c).

(c) That by a suitable choice of constants Z,

$$\left.\begin{aligned} D_F'&=Z_3D_{F1}'(e_1)\\ \Delta_F'&=Z_2\Delta_{F1}'(e_1)\\ \Gamma_\mu&=Z_1{}^{-1}\Gamma_{\mu 1}(e_1) \quad \text{and} \quad C_{\mu\nu}=Z_4{}^{-1}C_{\mu\nu 1}(e_1)\end{aligned}\right\} \quad (4)$$

where the renormalized (and the observed) charge of the meson

$$e_1=Z_1{}^{-1}Z_2Z_3{}^{\frac{1}{2}}e=Z_4{}^{-\frac{1}{2}}Z_2{}^{\frac{1}{2}}Z_3{}^{\frac{1}{2}}e'.$$

[5] The external lines of any proper part contain no self-energy insertions. In general a proper part is one which cannot be divided into two pieces joined by a single line.

[6] The factor for a 4-vertex is $-e^2\delta_{\mu\nu}$. For $\theta_{\mu\nu}$ as defined here we obtain the e-factor $(-1)^r e^{2r}e^{2s}$ where r is the number of 4-vertices and $2s$ that of 3-vertices, and then divide by $-e^2$. This gives $C_{\mu\nu}$ (to replace $\delta_{\mu\nu}$) as $=\delta_{\mu\nu}+\theta_{\mu\nu}$.

[7] This separation into convergent and divergent parts was given by Dyson (private communication).

[8] F. J. Dyson, Phys. Rev. 75, 486 (1949); P. T. Matthews, Phil. Mag. 185, XLI (1950).

[9] The necessity of this condition was pointed out by P. T. Matthews (reference 2).

(d) If $H_0(x)$ is gauge-invariant, so that $e=e'$, it has to be shown that the Z's defined satisfy a further condition,

$$Z_4^{-1}=Z_2Z_1^{-2}.$$

D_F', Δ_F', Γ_μ, $C_{\mu\nu}$ are the functions already defined, except that they now include the graphs with the new 2- and 4-vertices, as well as the original graphs. $C_{\mu\nu 1}=\delta_{\mu\nu}+\theta_{\mu\nu c}$ where $\theta_{\mu\nu c}$ is the function arising by adding together the (absolutely) convergent parts of each integral corresponding to each "original" C part. $\Gamma_{\mu 1}$, Δ_{F1}', D_{F1}' are defined in a similar manner from sums of convergent parts of the original integrals concerned. The procedure for obtaining the (absolutely) convergent part of an integral will be presently explained. It is to be noticed that the "convergent functions," on the right-hand side of the relations (4) appear to be expressed as functions of the renormalized charge e_1.

The general procedure for isolating divergences[10] from an n-fold integral was given in reference 2 in Sec. III. In this procedure the concept of true divergence plays an important part. In general if we subtract from any divergent integral over the variables $t_1\cdots t_n$, the true divergence over t_1 subintegration $D(t_1)\times$ the reduced integral (R) over $t_2t_3\cdots t_n+$ the true divergence over $t_2\times$ the reduced integral over $t_1t_3\cdots t_n+\cdots+$ the true divergence over $t_1t_2\times$ the reduced integral over $t_3\cdots t_n$ $+\cdots+\cdots+\cdots+$ finally the true divergence over $t_1t_2\cdots t_n$, the remainder is an integral which is absolutely convergent and is the convergent part of the n-fold divergent integral we started with. Each reduced integral corresponds to a graph obtained from the graph under consideration by omitting that part of the graph whose true divergence multiplies this particular reduced integral. If a subintegration is "superficially convergent," its true divergence is zero; while the true divergence over a subintegration $t_a\cdots t_b\cdots t_i\cdots t_jt_p\cdots t_q\cdots t_y\cdots t_z$ where $(t_a\cdots t_b)$, $(t_i\cdots t_j)$, $(t_p\cdots t_q)\cdots$ are k groups of variables belonging to nonoverlapping parts of the graph, equals $(-i)^{k-1}$ times the product of the true divergences over each of the k groups, $(t_a\cdots t_b)$, $(t_i\cdots t_j)$, $\cdots$ (Remark b, I, Sec. III[11]). It is in terms of the convergent parts of the integrals defined as above that the functions $C_{\mu\nu 1}$, $\Gamma_{\mu 1}$, etc. are defined. The true divergence of a subintegration corresponding to a meson self-energy graph is characterized by two divergent constants A, B; while all other true divergences are characterized by one divergent constant. The factors multiplying these constants $[(p^2+\kappa^2)$, for example, which occur multiplying B in the meson self-energy case] are absorbed in the reduced integral.

Considering the "original" graphs, the finite and physically significant expressions $C_{\mu\nu 1}$, $\Gamma_{\mu 1}\cdots$ are defined, as already stated, in the first place by an apparently arbitrary dropping (subtracting off) of the infinite terms (true divergences $\times$ the reduced integrals) from the infinite expressions $C_{\mu\nu}$, Γ_μ, etc. Some of the divergent terms thus subtracted can be interpreted under (a) and (b) as direct cancellations with terms from $\delta\kappa^2$ and $\delta\lambda$ arising in graphs which are not original so that, as we shall show, both these divergences as well as the "non-original" graphs need never be considered. The establishment of (c) and (d) shows that the remaining divergent terms isolated can equally well be interpreted as the extraction of infinite constant multiplicative factors Z, from $C_{\mu\nu}$, Γ_μ, $\cdots$, so that each of these functions, instead of appearing as a sum of infinite $(D\times R)+$finite terms, now appears, as in Eq. (4), as a product of a (divergent) constant (Z) multiplying finite and physically significant terms. After this is accomplished, these Z factors can be completely absorbed in renormalizing the charge, so that all infinities occurring in the theory can be eliminated.

The main difficulty of the proof lies in establishing (e). In each of Eqs. (4) the functions appearing on the left-hand side are completely known, while so are the finite parts of these functions $C_{\mu\nu 1}$, $\Gamma_{\mu 1}$, $\cdots$. Let us assume that the relations (4) hold, with the factors Z for the present unknown. Consider an irreducible C graph, T_{IRR}, for example, such that in its lines and vertices, self-energy, vertex, and C parts can be inserted without any one of these insertions overlapping with any other insertion. Each insertion is thus completely localizable and does not simultaneously modify more than one vertex or line. The result of these insertions analytically is that we replace in the integral for the irreducible graph Δ_F by Δ_F', D_F by D_F', $(p+p')_\mu$ by Γ_μ and $\delta_{\mu\nu}$ by $C_{\mu\nu}$. It was shown (Remark c, I, Sec. III), that if, for example, a meson line p has a self-energy insertion with associated momenta $t_at_b\cdots t_k$, and if this self-energy insertion defines no overlap with any other part of the graph, then we can arrange our general subtraction procedure so that we isolate divergences corresponding to the subintegrations t_a, $t_at_b\cdots$, etc., first. After the true divergence corresponding to $t_at_b\cdots t_k$, itself is removed, the integration over this set is left absolutely convergent and it can be performed unambiguously so that the variables t_a, t_b, $\cdots t_k$ no longer need be considered explicitly from this stage onward. Since this procedure of removing divergent terms corresponding to this self-energy insertion is precisely the one we would adopt for obtaining Δ_{F1}' from Δ_F', and, since we have assumed that, in the latter case (Eq. (4)), the divergent terms separated are

[10] The proof given by Rohrlich for the possibility of renormalization (reference 2) for scalar electrodynamics is invalidated because the general procedure for isolating divergences was not available. A "hierarchy" of divergences was somewhat arbitrarily defined and a prescription given according to which a divergence higher in the hierachy should be removed first. This procedure is not equivalent to the one given above and when overlaps occur, does not leave behind an absolutely convergent integral after the proposed subtractions. In a note added in proof (reference 2) Rohrlich has observed that the problem of "b-divergences" (overlaps) has not received proper treatment in his paper. The present paper deals with these difficulties.

[11] Abdus Salam, Phys. Rev. 82, 217 (1951), hereafter referred to as I.

completely interpreted as the extraction of the divergent factor Z_2 in $\Delta_F' = Z_2\Delta_{F1}'(e_1)$, we can make this same replacement for Δ_F' for this internal line p in the C graph under consideration. It is emphasized again that this is true if and only if no overlap occurs.

Going back to consider T_{IRR}, by inserting all S, V, and C parts in its lines and vertices we obtain an entire class of C parts $T_{\mu\nu}$ "derived" from T_{IRR}. Analytically we replace in the integral for T_{IRR}, Δ_F by $Z_2\Delta_{F1}'(e_1)$, D_F by $Z_3D_{F1}'(e_1)$, $(p+p')_\mu$ by $Z_1^{-1}\Gamma_{\mu 1}$, and $\delta_{\mu\nu}$ by $Z_4^{-1}C_{\mu\nu 1}$.

A C part of order $e^{2s+2r-2}$ with s 4-vertices and $2r$ 3-vertices contains $2r+s-1$ meson and $r+s-1$ photon lines. The foregoing replacements give an e factor

$$e^{2r+2s-2}Z_4^{-s}Z_1^{-2r}Z_2^{2r+s-1}Z_3^{r+s-1} = Z_4^{-1}e_1^{2s+2r-2}.$$

Δ_{F1}', D_{F1}', $\cdots$ are themselves power series in e_1, and their behavior for large p is precisely the same as that of Δ_F, D_F, etc. (D II Sec. VII), so that the new integral for $T_{\mu\nu}$ is again logarithmically divergent. Making a separation of the finite and the infinite part,

$$T_{\mu\nu} = Z_4^{-1}(T_d(e_1)\delta_{\mu\nu} + T_{\mu\nu c}(e_1)). \tag{5}$$

Equation (5) gives in a compact form the sum of inte-

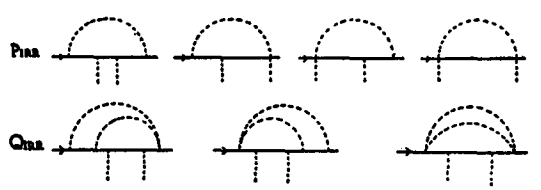


FIG. 1.

grals from the entire class of graphs $T_{\mu\nu}$, "derived" (uniquely) from the one graph T_{IRR} by insertions. This form for $T_{\mu\nu}$ will be referred to as the fundamental form. It is to be emphasized that the content of $T_{\mu\nu}$ is not altered in the fundamental form by expressing it thus in terms of the renormalized charge; and that the equality of both sides is exact considered up to any order in e_1 and e (expressed in terms of e_1).

If it were possible to make unambiguous nonoverlapping insertions in all the irreducible C parts, this procedure would have given us all the reducible and irreducible C parts. Analytically from each T_{IRR} we would have obtained $T_{\mu\nu}$ in the form (5) so that summing over all C parts and adding $\delta_{\mu\nu}$ for the 4-vertex itself we would obtain

$$C_{\mu\nu} = \delta_{\mu\nu} + Z_4^{-1}(R_d\delta_{\mu\nu} + \theta_{\mu\nu c}(e_1)).$$

But if Eqs. (4) hold, $C_{\mu\nu}$ ought to equal $Z_4^{-1}C_{\mu\nu 1}$. Thus

$$Z_4^{-1}C_{\mu\nu 1} = Z_4^{-1}(\delta_{\mu\nu} + \theta_{\mu\nu c}(e_1)).$$

This gives us an equation with just one unknown, Z_4, which could then be determined (and so the consistency of the relations (4) established.

However, as soon as overlaps occur the whole idea of an unambiguous insertion in a line or a vertex loses validity. The reduction of a reducible graph cannot be defined unambiguously and, conversely, the replacement of all vertex and line factors of irreducible parts by $C_{\mu\nu}$, Γ_μ, etc., leads to the counting of certain graphs more than once. [This is a definite redundance, quite apart from the "weight factors" occurring under rule (vi).] Thus, a proof on the aforementioned lines can no longer be given.

For C parts it is always possible to make unambiguous insertions in all the lines and 3-vertices. However there are certain classes of C parts in which a C part inserted in one 4-vertex appears simultaneously as a C part inserted at some other vertex.

For vertex parts the complexity of overlaps increases greatly and $C-V$ overlaps can occur. For self-energies, besides these overlaps, vertex parts overlap with vertex parts. Also, if any meson (or photon) line is opened in a self-energy graph, this leads to parts with four external lines which (except for photon-photon scattering graphs) always diverge. M parts, which have not been considered so far, can complicate the picture still further by producing simultaneously $M-M$, $M-C$, or $M-S$ overlaps.

A great simplification can be effected by applying a powerful technique first introduced by Ward[12] in spinor electrodynamics. Ward's technique can be extended to reduce the complexity of the overlaps to be considered. However, if a primitive divergent is logarithmically divergent and still suffers from overlaps, this technique in general fails. The general procedure to be followed then for obtaining a proof of (4) and for the construction of the relevant functions is that of categorization, as developed in I, Sec. IV. By analyzing the overlaps, such categories of reducible graphs are defined in which certain insertions of S, V, and C parts cause no overlap, so that the substitutions $C_{\mu\nu}$, Γ_μ, etc., can be made in a defined way. By considering the implications of the subtraction rules, I, Sec. III, it is then possible to derive recurrence relations for such graphs in terms of graphs of lower order in the same category and their true divergences. The required values for the Z's are obtained by substituting Eqs. (4) into these recurrence relations and demanding that they lead back to the fundamental form for the type of part under consideration. These various techniques are illustrated in Secs. 2, 3, and 4 and in Appendix I where it is assumed that condition (b) can be satisfied. A proof of this is given in Sec. 5.

2. C PARTS

The most general C part consists of an open polygon formed by the meson line entering (and leaving) the graph, this line (the base line of the graph) being joined by photon lines to one or more (possibly interconnected)

[12] J. C. Ward, Proc. Phys. Soc. (London) 64, 54 (1951). The author is deeply indebted to J. C. Ward for sending him a copy of his work prior to publication.

closed loops of meson lines with a structure inside them. If the external photon lines (with polarization vectors μ and ν) belong to the base line, the graph will be called a base-line graph. For base-line C parts it is possible to distinguish between $\theta_{\mu\nu}$ and $\theta_{\nu\mu}$. For such graphs $\theta_{\mu\nu}(p, p', q)$ is defined such that μ occurs topologically "before" ν and carries momentum q.

In order to construct the function $C_{\mu\nu}$ (and the corresponding convergent function $C_{\mu\nu 1}$) it is possible as explained in Sec. I, to take irreducible C parts and to make unambiguous insertions in their lines and 3-vertices, replacing in the corresponding integrals Δ_F by Δ_F', D_F by D_F', and $(p+p')_\mu$ by Γ_μ. Furthermore, it is possible to replace the factor $\delta_{\mu\nu}$ by $C_{\mu\nu}$ for the 4-vertices of all irreducible C parts except for the 4-vertices in the graphs P_{IRR} and Q_{IRR} shown in Fig. 1. (There is another class of graphs R which are defined in Appendix II and in the 4-vertices of which the foregoing insertion is also not valid.) Taking all other irreducible C parts and making the foregoing insertions gives all reducible C parts "derived" from them. If $T_{\mu\nu}$ denotes the sum of corresponding integrals, by counting the number of lines and vertices in the irreducible C parts in which insertions are being made, we have

$$T_{\mu\nu}=Z_4^{-1}[T_d(e_1)\delta_{\mu\nu}+T_{\mu\nu c}(e_1)], \tag{6}$$

where $T_d(e_1)$ is the sum of the true divergences from all such graphs.

In this section the graphs P_{IRR} and a linear chain of graphs "derived" from them (the whole class being called P) will be treated in detail. For Q and R we assume the result in Appendix II, namely, that with a proper choice of Z_4 the corresponding functions $Q_{\mu\nu}$ and $R_{\mu\nu}$ can also be expressed in the fundamental form of Eq. (6). In fact, at this stage $T_{\mu\nu}$ will be understood to contain $Q_{\mu\nu}$ and $R_{\mu\nu}$.

Consider the graph in Fig. 2, of order e.[7] We can obtain (2^5-1) other topologically distinct graphs from it by shrinking $1\to1$, $2\to2$, $3\to3$, $4\to4$, $5\to5$ in all possible ways. If $4\to4$, for example, the resulting 4-vertex will be (uniquely) numbered 4. In Fig. 2 the letters $a, b, \cdots$ give the polarization of the photon. The order of the numbers 1, 2, $3\cdots$ and the letters $a, b, c, \cdots$ are important. If 2 is a 4-vertex, it could be denoted equally well by δ_{ab} or by δ_{ba} so that the number of formally different graphs obtained from the graph in Fig. 2 by joining $1\to1$, $2\to2$, $\cdots$ is 3^5-1, when the graphs with factors δ_{ab} and δ_{ba} are treated as different.

Let $P_{\mu\nu}$ stand for the graphs obtained by drawing chain graphs of the type illustrated in Fig. 2 in all orders in the powers of e, and also the graphs obtained by joining $1\to1$, $2\to2$, $\cdots$, in all possible ways.

Let $[n]$ stand for such graphs (3^n in number) of order e^{2n}, $[n|1]$ stand for all graphs $[n]$ such that the first vertex is necessarily a 4-vertex. These are 3^n in number. Similarly, let $[n|i,j]$ stand for all graphs(3^{n-1} in number) which have their ith and jth vertices necessarily δ_{pq} and δ_{rs} 4-vertices (the order of photon polarization-vectors p, q and r, s is to be noted), irrespective of the character of the other vertices, and so on for $[n|i,j,\cdots]$. Let $X_{\mu\nu}$ stand for all those graphs of all orders (and the sum of corresponding integrals) from among the P graphs which have their last end vertex a $\delta_{s\nu}$ 4-vertex, irrespective of the character of other vertices, $Y_{\mu\nu}$ stand for all graphs with the first and the last end-vertices, $\delta_{\mu a}$ and $\delta_{\mu a}$ vertices, and $Z_{\mu\nu}$ for those with the first end-vertex a 4-vertex $\delta_{\mu a}$, again irrespective of the character of the other vertices. Thus, Y, for example, equals

$$[1|1,2]+[2|1,3]+\cdots+\cdots+[n|1,n+1]+\cdots.$$

With the foregoing definitions we prove Lemma 1.

Lemma 1

$$P_{\mu\nu}=P_d(\delta_{\mu\tau}+X_{\mu\tau})(\delta_{\tau\rho}-P_dY_{\tau\rho}+P_d^2Y_{\tau\rho}{}^2-\cdots)(\delta_{\rho\nu}+Z_{\rho\nu})+P_{\mu\nu c}. \tag{7}$$

Here P_d is the sum of the true divergences from the graphs P. In particular if P_n stands for the sum of the true divergences from $[n]$, then $P_d=P_1+P_2+P_3+\cdots$

For the proof we make use of the property of a 4-vertex occurring in this type of linear chain of graphs P to split the chain into two parts.

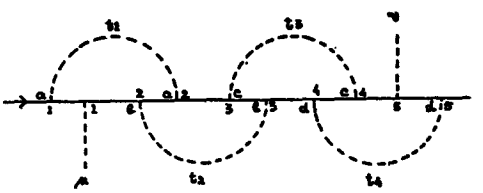


FIG. 2.

Consider, for example, the graphs in Fig. 3. It is obvious that the double integral over t_1t_2 corresponding to Fig. 3(a) is the product of two single integrals over t_1 and t_2, so that symbolically $(a)=(b)\times(c)$. In the sequel we shall further specialize this splitting off of graphs. A split will be made for a δ_{ab} vertex but it will never be made for a δ_{ba} 4-vertex.

$[n]$ is an n-fold integral over variables $t_1t_2\cdots t_n$, each one of the subintegrations being logarithmically divergent. If $ij\cdots q$ is a group of indices from the set $123\cdots n$ (where $ij\cdots$ run consecutively) then isolating the divergence over the subintegration $t_it_j\cdots t_q$ gives $P_{(q-i+1)}\times$the corresponding reduced integral. This reduced integral contains a 4-vertex of the type with a factor δ_{ab} (and not δ_{ba}) corresponding to the divergence which has been separated and this immediately splits the chain (if, of course, this 4-vertex is not one of the two possible end 4-vertices).

To obtain the coefficient of $(P_d)^f$ in (7), consider all subsets of indices $123\cdots n$, such that each subset divides naturally into f groups, the indices in each group running consecutively, the different groups being such that if k is the first and i the last index of two consecutive groups, then $k-i\geq1$. The true divergence for a subintegration corresponding to these subsets consists of $(-1)^{f-1}P_\alpha P_\beta\cdots(f$ factors), (I, Remark b, Sec. III) corresponding to the number (f) of nonoverlapping graphs, the true divergences of which are in fact thus being simultaneously removed. The reduced integral itself necessarily contains f 4-vertices (again of the type δ_{ab} and not δ_{ba}), which may split the graph according to the scheme XY^{f-1}, Y^{f-1}, $Y^{f-1}Z$, or $XY^{f-1}Z$. By considering all possible subsets and the groups in them of the type mentioned we readily establish the lemma.

We can show this by considering, as an example, the coefficient of P_d^3 in (7) in detail. Only graphs up to $[n]$ are considered. In our symbolic notation $[1|1,2]\times[n-1|1]$ would be equal to $[n|1,2]$. We consider all subsets of indices which fall into two nonoverlapping groups. If the first of these two groups starts with index 1, we have the following terms to separate:[13]

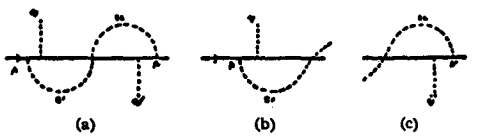
$$
\begin{aligned}
&-P_1[1|1,2]\times \\
&\quad\Big[P_1[n-3|1]+P_2[n-4|1]+P_3[n-5|1]+\cdots+P_{n-2} \\
&\qquad +P_1[n-4|1]+P_2[n-5|1]+\cdots+P_{n-3} \\
&\qquad\qquad +P_1[n-5|1]+\cdots+P_{n-4} \\
&\qquad\qquad\qquad +\cdots+\cdots \\
&\qquad\qquad\qquad\quad +P_1\Big] \\
&-P_2[1|1,2]\times \\
&\quad\Big[P_1[n-4|1]+P_2[n-5|1]+\cdots+P_{n-3} \\
&\qquad +P_1[n-5|1]+\cdots+P_{n-4} \\
&\qquad\qquad +\cdots+\cdots \\
&\qquad\qquad\quad +P_1\Big] \\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&-P_{n-2}[1|1,2]P_1
\end{aligned}
\qquad (8)
$$

The terms (8) sum up to

$$
\begin{aligned}
&=-P_d[1|1,2]P_d(1+[1|1]+[2|1]+[3|1]+\cdots+[n-3|1]) \\
&=-P_d^2[1|1,2](\delta_{\mu\nu}+Z_{\mu\nu})
\end{aligned}
$$

if the equality of both sides is supposed to hold to the order e^{2n}. Also, there are terms such as

$$
\begin{aligned}
&-P_1[2|1,3]\Big[P_1[n-4|1]+P_2[n-5|1]+\cdots+P_{n-3} \\
&\qquad +P_1[n-5|1]+\cdots+P_{n-4} \\
&\qquad\qquad +\cdots+\cdots \\
&\qquad\qquad\quad +P_1\Big] \\
&-P_2[2|1,3]\Big[\qquad P_1[n-5|1]+\cdots\cdots\cdots\Big] \\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$



(a) (b) (c)

FIG. 3.

which sum up to $-P_d^3[2|1,3](\delta_{\mu\nu}+Z_{\mu\nu})$. We similarly obtain $-P_d^3[3|1,4](\delta_{\mu\nu}+Z_{\mu\nu})$ and so on, concluding with $-P_1^3[n-2|1,n-1]$. The sum of all these terms is $-P_d^3Y_{\mu\rho}(\delta_{\rho\nu}+Z_{\rho\nu})$.

If we consider the two groups of indices such that the first group does not contain the index 1, it may be easily verified that the sum of corresponding contributions is $-P_d^3X_{\mu\tau}Y_{\tau\rho}(\delta_{\rho\nu}+Z_{\rho\nu})$ so that the entire sum is $-P_d^3(\delta_{\mu\tau}+X_{\mu\tau})Y_{\tau\rho}(\delta_{\rho\nu}+Z_{\rho\nu})$ precisely as in (7).

We now notice an important property of these graphs P. Whereas the insertion of any graph P_{ab} for δ_{ab} leads only to a graph already belonging to P, an insertion of P_{ba} for δ_{ba} gives a new graph unambiguously.

In order to obtain all reducible graphs "derived" from the graphs in Fig. 1. We proceed as follows:

(i) Draw all graphs P of all orders in e^2.

(ii) Insert at all 4-vertices δ_{ji}, all reducible and irreducible C parts including P^*_{ji}; analytically, replace δ_{ji} in the integrals by C_{ji}.

(iii) Insert at all 4-vertices δ_{ij}, the graphs T_{ij}; i.e., replace δ_{ij} by $\delta_{ij}+T_{ij}$. In the lines and the 3-vertices, all self-energy and vertex parts are inserted. The result after these insertions is $P_{\mu\nu}{}^*$.

It is easy to verify that no graph occurs more than once provided we adhere to the convention of distinguishing between δ_{ij} and δ_{ji} and further that all graphs "derived" from the graphs P_{IRR} do appear in the foregoing scheme of categorization, so that

$$C_{\mu\nu}=\delta_{\mu\nu}+T_{\mu\nu}+P_{\mu\nu}{}^*. \qquad (9)$$

Since all the insertions (i) to (iii) are unambiguous and localizable, we can replace Δ_F by $\Delta_F'=Z_2\Delta_{F1}'$, D_F by $D_F'=Z_3D_{F1}'(e_1)$, $(p+p')_\mu$ by $Z_1^{-1}\Gamma_{\mu1}(e_1)$, δ_{ji} by $C_{ij}=Z_4^{-1}C_{ji1}(e_1)$ while δ_{ij} is replaced by

$$
\begin{aligned}
\delta_{ij}+T_{ij}&=\delta_{ij}+Z_4^{-1}(T_d\delta_{ij}+T_{ijc}(e_1)) \\
&=Z_4^{-1}[(T_d+Z_4-1)\delta_{ij}+\delta_{ij}+T_{ijc}(e_1)] \\
&=Z_4^{-1}[S\delta_{ij}+S_{ij1}(e_1)].
\end{aligned}
$$

(definition of S)

The effect of these substitutions is to give Z_4^{-1} times the correct number of Z factors to renormalize the charge while all integrals now appear explicitly as functions of e_1. While at each δ_{ji} vertex, δ_{ji} has been replaced by C_{ji1}, each δ_{ij} vertex gives rise to two terms; one in which the integral corresponding to the graph is merely multiplied by S, the other in which δ_{ij} is replaced by $S_{ij1}(e_1)$.

Denote by $P_{\mu\nu}{}^\times$ the sum of integrals corresponding to all graphs thus formed, for whose 4-vertices appear the factors S_{ij1} or C_{ji1} while for the lines and 3-vertices Δ_{F1}, D_{F1}, $\Gamma_{\mu1}$ appear. $X_{\mu\nu}{}^\times$ denotes the corresponding sum for all graphs in which only the last end-vertex is an unchanged 4-vertex δ_{sr} while all other 4-vertices are replaced by factors S_{ij1} or C_{ji1}. $Y_{\mu\nu}{}^\times$ similarly denotes the sum of integrals for all graphs with only the first and the last end-vertices unchanged 4-vertices, and $Z_{\mu\nu}{}^\times$ the sum for graphs with only the first end-vertex an unchanged 4-vertex.

With these definitions we prove the following lemma.

Lemma 2

$$
\begin{aligned}
P_{\mu\nu}{}^*=Z_4^{-1}[&S(\delta_{\mu\tau}+X_{\mu\tau}{}^\times)(\delta_{\tau\rho}+SY_{\tau\rho}{}^\times \\
&+S^2Y_{\tau\rho}{}^{\times2}+\cdots)(\delta_{\rho\nu}+Z_{\rho1}{}^\times)-S\delta_{\mu\nu}+P_{\mu\nu}{}^\times].
\end{aligned}
\qquad (10)
$$

Defining $[n|l,m]^\times$, etc., to represent all graphs $[n|l,m]$ for the lines and vertices of which replacements Δ_{F1}', D_{F1}', $\Gamma_{\mu1}$, C_{ji1} or S_{ij1} have been made, except at the lth and mth 4-vertices where the factors δ_{pq} and δ_{rs} remain unchanged, the proof of the lemma follows by noticing that

$$
\begin{aligned}
[n]^*=[n]^\times+&S([n|1]^\times+[n|2]^\times+\cdots) \\
&+S^2([n|1,2]^\times+[n|1,3]^\times+\cdots \\
&+[n|i,l]^\times+\cdots)+S^3([n|1,2,3]^\times+\cdots \\
&+[n|i,l,k]^\times+\cdots)+\cdots+\cdots \\
&+S^{n+1}[n|1,2,3,\cdots,n+1]^\times.
\end{aligned}
\qquad (11)
$$

[13] To verify that a term like $-P_1[1|1,2]P_3[n-5|1]$, for example, takes account of all the 3^{n+1} graphs in $[n]$, one can make a check as follows: $[n-5|1]$ contains 3^{n-5} graphs, P_3 is the true divergence of 3^{n+1} graphs in $[3]$, $[1|1,2]$ accounts for 3^0 and P_1 for 3^3 graphs. Thus, the foregoing term expresses the sum of relevant contributions from all the 3^{n+1} graphs in $[n]$.

By splitting the graphs and arranging the sumations as in Lemma 1, the result (10) is established.

Since Δ_{F1}', D_{F1}', etc., have the same behavior for large values of p, as the corresponding functions Δ_F, D_F (DII, Sec. VII), an immediate consequence of Lemma 1 is Lemma 3.

Lemma 3

$$P_{\mu\nu}{}^{\times}=P_d{}^{\times}(\delta_{\mu\tau}+X_{\mu\tau}{}^{\times})(\delta_{\tau\rho}-P_d{}^{\times}Y_{\tau\rho}{}^{\times}+P_d{}^{\times 2}Y_{\tau\rho}{}^{\times 2}-\cdots)(\delta_{\rho\nu}+Z_{\rho\nu}{}^{\times})+P_{\mu\nu c}{}^{\times}. \quad (12)$$

We have now reached the stage in our inductive argument, when we can establish the consistency of our procedure by an explicit choice of the unknown constant Z_4. We desire to choose Z_4 such that

$$C_{\mu\nu}=\delta_{\mu\nu}+T_{\mu\nu}+P_{\mu\nu}{}^{*}=Z_4{}^{-1}C_{\mu\nu 1}=Z_4{}^{-1}(\delta_{\mu\nu}+T_{\mu\nu c}+P_{\mu\nu c}{}^{\times}). \quad (13)$$

From (6), (11), and (12), however,

$$\delta_{\mu\nu}+T_{\mu\nu}+P_{\mu\nu}{}^{*}=\delta_{\mu\nu}+Z_4{}^{-1}[(S-T_d)\delta_{\mu\nu}+P_{\mu\nu c}{}^{\times}+T_{\mu\nu c}]+Z_4{}^{-1}(\delta_{\mu\tau}+X_{\mu\tau}{}^{\times})[S/(1-SY^{\times})+P_d{}^{\times}/(1+P_d{}^{\times}Y^{\times})](\delta_{\rho\nu}+Z_{\rho\nu}{}^{\times}). \quad (14)$$

Let $S+P_d{}^{\times}=Z_4+T_d+P_d{}^{\times}-1=0$, then $[S/(1-SY^{\times})+P_d{}^{\times}/(1-P_d{}^{\times}Y^{\times})]\equiv 0$ and simultaneously Eq. (13) is satisfied.

Thus, with the choice $Z_4=1-T_d-P_d{}^{\times}$ we finally establish that we can express $C_{\mu\nu}=Z_4{}^{-1}C_{\mu\nu 1}(e_1)$. Furthermore, by thus expressing Z_4 in terms of the true divergences of all original C parts, we have also expressed $P_{\mu\nu}{}^{*}$ in Lemma 2 in the fundamental form

$$P_{\mu\nu}{}^{*}=Z_4{}^{-1}[P_d{}^{\times}(e_1)\delta_{\mu\nu}+P_{\mu\nu c}{}^{\times}(e_1)]. \quad (15)$$

3. VERTEX PARTS AND MESON SELF-ENERGY GRAPHS

We now desire to show that it is possible to choose the constants Z_1 and Z_2 such that $\Delta_F'=Z_2\Delta_{F1}'$ and $\Gamma_\mu=Z_1{}^{-1}\Gamma_{\mu 1}$ and furthermore that $Z_4{}^{-1}=Z_2Z_1{}^{-2}$. The proof can be made to depend on that in Sec. II for C parts, by employing a technique due to Ward. It will be found that at least for Δ_{F1}', it is not even necessary to obtain its value by evaluating the absolutely convergent parts of all self-energy integrals. For this purpose we utilize the following differential identities,[14]

$$\Lambda_\mu(p,p)=-(1/2\pi i)(\partial/\partial p_\mu)\Sigma^{*} \quad (16)$$

$$(\partial/\partial p+\partial/\partial p')_\nu\Lambda_\mu(p,p')=\theta_{\mu\nu}(p,p',p'-p)+\theta_{\nu\mu}(p,p',0). \quad (17)$$

To prove these identities, we notice the differential relation,

$$-(1/2\pi i)(\partial/\partial p_\mu)\Delta_F(p)=\Delta_F(p)2p_\mu\Delta_F(p),$$

which correctly describes the insertion of an external photon line (with its energy-momentum set equal to zero) in a meson line with momentum p. A second differentiation with respect to ν describes not only the insertion of another photon 3-vertex on the same meson line, but also the complication of the first 3-vertex into a 4-vertex, with the proper weight factor.

To prove (16) we notice first that the sum of the contributions to $\Lambda_\mu(p,p)$ from all (reducible or irreducible) non-base-line vertex parts vanishes identically. Let the momentum p be always associated with the base line. If t_i represent the momentum variables associated with closed meson loops (connected to the base line and to each other), then the sum of the contributions to $\Lambda_\mu(p,p)$ from non-base-line vertex parts is

$$\sum_i\int(\partial/\partial t_i)[F_\mu(p,t_a)]dt_a$$

which vanishes identically because F_μ is uniformly small for large values of momenta t_a.

Similarly, for a C part, if the photon ν has its energy-momentum zero, and does not belong to the base line, the sum of the contributions for such graphs to $\theta_{\mu\nu}(p,p',p'-p)$ and $\theta_{\nu\mu}(p,p',0)$ identically vanishes.

Thus, (a) the only vertex parts giving a contribution to $\Lambda_\mu(p,p)$ are the base-line vertex parts,

(b) The only C parts contributing to $\theta_{\mu\nu}(p,p',p'-p)+\theta_{\nu\mu}(p,p',0)$ are those for which ν belongs to the base line.

If now we associate p with the base line for meson self-energy graphs, a differentiation gives precisely all the base-line vertex parts and no photon line is differentiated. This establishes (16).

Similarly on associating p, p' with the base line in vertex parts, the operator $[\partial/\partial p+\partial/\partial p']$ would give all the C parts with ν belonging to the base line and these are the only ones contributing to the right-hand side of (17), as shown. If μ itself does not belong to the base line, but to one of the closed meson loops, some photon line joining this loop to the base line must carry momentum $(p-p'+t_a)$, while a part of the loop itself may have momentum variables $(p-p'+t_b)$. The operator $[\partial/\partial p+\partial/\partial p']$ insures that neither this photon line, nor any part of the closed loop is differentiated. This establishes (17).

The foregoing proofs depend on a very particular choice of momenta; p, p' must be associated with the base line. This choice can sometimes lead to difficulties in isolating divergences. This question will be examined further in Sec. V, when we deal with M parts.

We now show by an inductive application of the identities (16) and (17) that these differential relations give rise to relations between true divergences and also hold for the convergent parts of the functions concerned as well. These convergent parts are the same parts of the divergent integrals as are obtained by our subtraction procedure.

[14] Abdus Salam, Phys. Rev. 79, 910 (1950). These (or similar) identities were derived independently by F. J. Dyson (private communication), F. Rohrlich (reference 2). Rohrlich's derivation of the relations between divergent constants from them was, however, incomplete because difficulties connected with overlaps were not noticed.

We treat the case of vertex parts in detail. It is obvious that the C parts produced by applying the operator $\partial/\partial p+\partial/\partial p'$ to an irreducible vertex part are themselves irreducible. (An extra external line always improves the chances of irreducibility and reduces overlaps.) Thus, from (2), for irreducible graphs,

$$2L\delta_{\mu\nu}+(\partial/\partial p+\partial/\partial p')_\nu\Lambda_{\mu c}(p,\,p') =2R\delta_{\mu\nu}+\theta_{\mu\nu c}(p,\,p',\,p'-p)+\theta_{\nu\mu c}(p,\,p',\,0). \quad (18)$$

Setting $p=p'=p_0$ shows that $L=R$ and also that the differential relation holds for convergent parts as well.

Consider now a reducible vertex part. The integral corresponding to it is $\Lambda_\mu=\Lambda_{\mu c}+$the true divergence constant $L\times(p+p')_\mu+$true divergent constants $D\times$the reduced integrals, where each reduced integral represents a vertex part of a lower order. Therefore $(\partial/\partial p+\partial/\partial p')_\nu\Lambda_\mu=(\partial/\partial p+\partial/\partial p')_\nu\Lambda_{\mu c}+$divergent constants $D\times(\partial/\partial p+\partial/\partial p')_\nu$ applied to the reduced integrals.

Consider now the C parts produced by this operator from the given vertex part. We can show that for such graphs $\theta_{\mu\nu}+\theta_{\nu\mu}=\theta_{\mu\nu c}+\theta_{\nu\mu c}+2\delta_{\mu\nu}\times$the true divergence constant $R+$the same true divergence constants D as above$\times$precisely such C parts as are obtained by

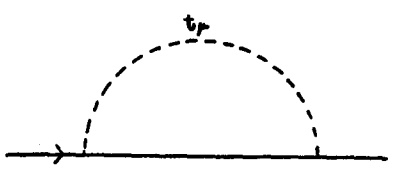


FIG. 4.

applying the differentiation operator to the aforementioned reduced vertex graphs.

A term by term comparison would thus allow us to write an equation of type (18), from which the equality of L and R and a differential relation between $\Lambda_{\mu c}$ and $\theta_{\mu\nu c}$ can be deduced.

The proof of this proceeds inductively. Since the differentiation operator affects only the base line in the given vertex part V, the true divergences for those parts of this graph which do not contain portions of the base line are the same as the true divergences for the relevant C parts obtained by differentiating V, while the reduced integrals in the two cases are themselves connected by the differential relation. Thus it is only for those subintegrations which extend over a part of the graph containing portions of the base line that a proof is needed. If, in the vertex part under consideration, such subintegrations correspond to inserted self-energy, vertex part or C part divergences, the differentiation operator converts them into, respectively, vertex part, or C part divergences, while the last type of divergence becomes superficially convergent. To bring out the points involved we consider, as an example, the case of a vertex part V, with its base line containing as an insertion the irreducible self-energy graph S (subintegration t_r) in Fig. 4.

Disregarding the mass renormalizing constant A from this self-energy graph S, the relevant divergent term is $B'\times V_{\text{Red}}$ where B' is charge-renormalizing constant from S, and V_{Red} is obtained from V by removing this self-energy insertion from the base line altogether.

Let us now suppose that the number of vertices on the base line of V is n, of which k are 3-vertices. The differentiation operator produces from the graph V, precisely $(n-1+2k)$ C parts. Of these 5 are such that they represent the replacing of S in the base line of V, by the 5 (irreducible) vertex parts which a differentiation of S yields, while in the remaining $(n-6+2k)$ C parts, the insertion S remains unchanged in the base line. Therefore the divergence separated corresponding to the subintegration t_r from this class of $(n-1+2k)$ C parts is $=B'\times C_{\text{Red}}'+L'\times C_{\text{Red}}''$. C_{Red}' are the $(n-6+2k)$ reduced C parts obtained by omitting S altogether from the base line, L' is the sum of the true divergent constants from the 5-vertex parts obtained by differentiating S, while C_{Red}'' is the reduced C part (just one graph) obtained by omitting any of the 5-vertex parts, and replacing it by a 3-vertex. Two of the graphs C_{Red}' are identical with the graph C_{Red}'', i.e., the reduced graphs from those two C graphs obtained by inserting the external photon line ν, just "before" and just "after" S, on the base line of V.

Now from Eq. (2) and (16), $B'=-L'$ (all graphs concerned in this relation are irreducible and the factor $1/2\pi i$ is absorbed in the definition of the reduced integral C_{Red}''); consequently, the term separated corresponding to the subintegration t_r is $=B'\times C_{\text{Red}}$ where C_{Red} represents $(n+2k-7)$ (different) C graphs; but these are precisely the graphs which we obtain by applying the differentiation operator $(\partial/\partial p+\partial/\partial p')$ to the graph V_{Red}, as can be checked by noticing that V_{Red} has $n-2$ vertices on the base line, of which $k-2$ are 3-vertices and therefore the number of C parts obtained by differentiating V_{Red} is $n-3+2(k-2)=n+2k-7$.

The proof is thus arranged, by considering each of the true divergence$\times$the reduced integral separated from the vertex part V. It is shown that for the corresponding C parts, the true divergences$\times$reduced integrals can be grouped such that the true divergences are equal by an inductive application of (16) and (17), and, furthermore, that the reduced integrals from C parts are precisely those which can be obtained by applying the differentiation operator to the reduced graphs. Since the argument proceeds by considering all subintegrations, overlaps are automatically taken care of.

The proof, therefore, of the required relations for a vertex part of order e^{2n+2} depends on establishing relations between true divergent constants of corresponding S, V, and C parts up to order e^{2n}. As the induction starts with irreducible graphs for which such relations are obviously true from the very definitions, the foregoing result follows. Thus, finally (putting $p=p=p_0$),

we have

$$(\partial/\partial p+\partial/\partial p')_\nu\Lambda_{\mu c}(p, p') = \theta_{\mu\nu c}(p, p', p'-p)+\theta_{\nu\mu c}(p, p', 0). \quad (19)$$

We shall now show that Z_1 can be chosen such that

$$\Gamma_\mu=(p+p')_\mu+\Lambda_\mu(p, p')=Z_1^{-1}[(p+p')_\mu+\Lambda_{\mu e}(p, p', e_1)]$$

where by definition $\Lambda_{\mu c}(p_0, p_0)=0$. Let

$$\theta_{\mu\nu}(p, p', p'-p)+\theta_{\nu\mu}(p, p', 0)=2G_{\mu\nu}(p'+p, p'-p).$$

Then defining

$$r^\lambda=(p+p')\lambda+2p_0(1-\lambda), \quad s^\lambda=(p'-p)\lambda$$

we rewrite (17) as

$$(\partial/\partial r_\nu^\lambda)\Lambda_\mu(r^\lambda, s^\lambda)=G_{\mu\nu}(r^\lambda, s^\lambda)$$

but

$$(\partial/\partial\lambda)\Lambda_\mu(r^\lambda, s^\lambda)=(p+p'-2p_0)_\nu G_{\mu\nu}(r^\lambda, s^\lambda) + (p'-p)_\nu\partial\Lambda_\mu/\partial s_\nu^\lambda;$$

consequently,

$$\Lambda_\mu(p, p')-\Lambda_\mu(p_0, p_0) = \int_0^1 d\lambda(p+p'-2p_0)_\nu G_{\mu\nu}(r^\lambda, s^\lambda) + \int_0^1 d\lambda(p'-p)_\nu\partial\Lambda_\mu/\partial s_\nu^\lambda. \quad (20)$$

Also, from (19),

$$\Lambda_{\mu c}(p, p')=\int_0^1 d\lambda(p+p'-2p_0)_\nu G_{\mu\nu c}(r^\lambda, s^\lambda) + \int_0^1 d\lambda(p'-p)_\nu\partial\Lambda_{\mu c}/\partial s_\nu^\lambda. \quad (21)$$

We have already shown, however, that Z_4 can be chosen such that

$$\theta_{\mu\nu}(p, p', q)=\delta_{\mu\nu}(Z_4^{-1}-1)+Z_4^{-1}\theta_{\mu\nu c}(p, p', q, e_1).$$

Therefore, (20) can be rewritten as

$$\left[(p+p')_\mu+\Lambda_\mu(p, p')-\int_0^1 d\lambda(p'-p)_\nu\partial/\partial s_\nu^\lambda(r_\mu^\lambda+\Lambda_\mu)\right] -[(2p_0)_\mu+\Lambda_\mu(p_0, p_0)]=Z_4^{-1}\left[(p+p')_\mu + \int_0^1 d\lambda(p+p'-2p_0)_\mu G_{\mu\nu c}\right]-Z_4^{-1}[(2p_0)_\mu]. \quad (22)$$

Since by definition, $\Gamma_{\mu 1}=(p+p')_\mu+\Lambda_{\mu c}(p, p', e_1)$, we have, from (22) using (21)

$$\Gamma_\mu(p, p')-\int_0^1 d\lambda(p'-p)_\nu\partial\Gamma_\mu/\partial s_\nu^\lambda = Z_4^{-1}\left[\Gamma_{\mu 1}(p, p')-\int^1 d\lambda(p'-p)_\nu\partial\Gamma_{\mu 1}/\partial s_\nu^\lambda\right]. \quad (23)$$

This leads us to infer the required result, namely, that $\Gamma_\mu=Z_4^{-1}\Gamma_{\mu 1}(e_1)$.

Knowing the value of $G_{\mu\nu c}$ we cannot obtain that of $\Lambda_{\mu c}(p, p')$ from Eq. (21), but that of $\Lambda_{\mu c}(p, p)$ can be readily found by a simple integration. Since

$$\sum_c{}^*(p)=-2\pi i\int_0^1 d\lambda(p-p_0)_\mu\Lambda_{\mu c}(p^\lambda, p^\lambda) \quad (24)$$

where $p^\lambda=p\lambda+p_0(1-\lambda)$, it is not needful for the derivation of $\sum_c{}^*(p)$ to consider the meson self-energy graphs any longer. The proof that $\Delta_F'=Z_4\Delta_{F1}'(e_1)$ follows from Eq. (16) in a similar manner to (23); in this case $\sum_M{}^*(p_0)$ is canceled by combining it with the contribution to the self-energy arising from the mass-renormalization term $-\delta\kappa^2\phi^*\phi$ in the Hamiltonian. We have also incidentally shown that $Z_1=Z_2=Z_4$, which satisfies condition (d) of the introduction.

4. THE FUNCTION D_F'

For photon self-energy graphs, we extend a formal technique introduced by Ward and define the functions, $\Delta_\mu(p)$ and $\Phi_{\mu\nu}(p)$, by equations,

$$-1/2\pi(\partial/\partial p_\mu)\Pi^*=\Delta_\mu(p) \quad (25)$$

$$(\partial/\partial p_\mu)\Delta_\mu(p)=\Phi_{\mu\nu}(p). \quad (26)$$

Also, let

$$W_\mu(p)=2ip_\mu+\Delta_\mu(p) \quad (27)$$

$$X_{\mu\nu}(p)=2i\bar{\delta}_{\mu\nu}+\Phi_{\mu\nu}(p). \quad (28)$$

The bar in $\bar{\delta}_{\mu\nu}$ distinguishes it from the $\delta_{\mu\nu}$ for a 4 C-vertex. These functions $W_\mu(p)$, and $X_{\mu\nu}$ stand in analogy to Γ_μ and $C_{\mu\nu}$ while Δ_μ and $\Phi_{\mu\nu}$ are analogous to Δ_μ and $\theta_{\mu\nu}$.

By integration we obtain from the foregoing

$$\Delta_\mu(p)=\int_0^1 d\lambda\Phi_{\mu\nu}(\lambda p)p_\nu \quad (29)$$

and

$$\Pi^*(p)=-2\pi\int_0^1 d\lambda\Delta_\mu(\lambda p)p_\mu. \quad (30)$$

Thus, in order to obtain $\Pi^*(p)$ we only need construct the function $\Phi_{\mu\nu}(p)$ which is logarithmically divergent. Dropping its divergent terms, the integrations in (29) and (30) give the desired convergent part of Π^*. In actual fact Π^* is not a scalar but a tensor[15] and its correct form (for the case of a simple irreducible photon self-energy part) is given by

$$\Pi_{\rho\kappa}{}^*=(\delta_{\rho\kappa}p^2-p_\rho p_\kappa)(C+D_c(p^2)).$$

The terms $-p_\rho p_k$ are omitted on account of charge conservation.[16]

[15] Julian Schwinger, Phys. Rev. 76, 790 (1949), Appendix.
[16] R. P. Feynman, Phys. Rev. 76, 781 (1949).

note $p_\rho(\delta_{\rho\kappa}p^2-p_\rho p_\kappa)\equiv 0$

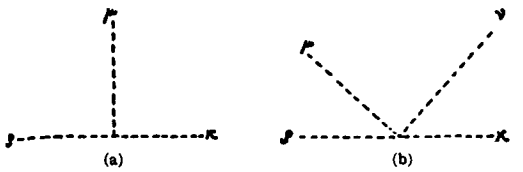


FIG. 5.

Thus, the function $\Phi_{\mu\nu}$ is in fact of the form:

$$\Phi_{\mu\nu,\,\rho\kappa}=[C+D_c(p^2)][2\bar{\delta}_{\mu\nu}\delta_{\rho\kappa}-\delta_{\rho\mu}\delta_{\kappa\nu}-\delta_{\rho\nu}\delta_{\kappa\mu}]+(\delta_{\rho\kappa}p^2-p_\rho p_\kappa)(\partial^2/\partial p_\mu\partial p_\nu)D_c(p^2).$$

However, if the terms $-\delta_{\rho\mu}\delta_{\kappa\nu}-\delta_{\rho\nu}\delta_{\kappa\mu}$ and $-p_\rho p_\kappa$ are dropped after the evaluation of $\Phi_{\mu\nu,\,\rho\kappa}$, we have

$$\Phi_{\mu\nu,\,\rho\kappa}=2\bar{\delta}_{\mu\nu}\delta_{\rho\kappa}(C+D_c(p^2))+\delta_{\rho\kappa}(\partial^2/\partial p_\mu\partial p_\nu)D_c(p^2)=\delta_{\rho\kappa}\Phi_{\mu\nu}.$$

In order to evaluate the function $\Phi_{\mu\nu}$ we must examine the implications of our formal differentiation $\partial/\partial p$. In general a photon self-energy graph may have both its external photon lines, μ and ν, belonging either to the same meson loop or to two distinct interconnected meson loops. In the latter case the momentum p must run along some photon lines joining the two distinct loops. The differentiation $\partial/\partial p$ which applied to a meson line carrying momentum p graphically signifies insertion of an external photon line (with energy-momentum zero), when applied to a photon line gives a new type of "vertex," illustrated in Fig. 5 with the vertex factor $2ip_\mu$ since $[-1/2\pi i(\partial/\partial p_\mu)D_F=D_F\cdot 2ip_\mu\cdot D_F]$. Thus to obtain the graphs corresponding to $\Phi_{\mu\nu}$ we first adopt some convention for the path of the variable p through the photon self-energy graphs. After "differentiating" all the photon self-energy graphs twice, we obtain the graphs corresponding to $\Phi_{\mu\nu}$ which contain vertices of the type in Fig. 5(a) and (b). The number and complexion of these graphs is governed by the convention[17] we adopt for the path of p but as the integrations (29) and (30) show, the choice of a convention does not matter as far as the evaluation of $\Pi^*(p)$ is concerned. After drawing all graphs $\Phi_{\mu\nu}$ we select the irreducible $\Phi_{\mu\nu}$, the criterion for irreducibility being the same as in Sec. 1, with the obvious extension that if a photon self-energy graph inside an internal photon line is differentiated, in order to obtain the irreducible skeleton, this is replaced by either of the vertices in Fig. 5 as the case may be. The advantage, as we shall see, in considering the function $\Phi_{\mu\nu}$ rather than Π^* itself is that the overlaps occurring for $\Phi_{\mu\nu}$ are relatively simple and it is easier to categorize the graphs corresponding to $\Phi_{\mu\nu}$ rather than Π^*. It may be emphasized that Δ_μ and $\Phi_{\mu\nu}$ are "fictitious" primitive divergents, being entirely defined by Eqs. (25) and (26) and have no real place in the theory. The convention for the path of p adopted here is illustrated in Fig. 6. Single arrows give the direction of charge, while double arrows follow p. In Fig. 6(a), p runs along the edge of the diagram; in (b) the path of p is "complementary," in the sense that if in (a), a portion of a closed loop is "differentiated," (b) insures the "differentiation" of the remaining portion of this same closed loop. Thus

$$(-1/2\pi)\partial/\partial p_\mu[(a)+(b)]=2\Delta_\mu(p).$$

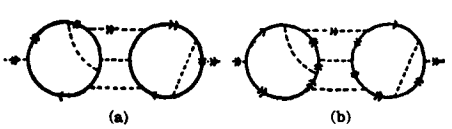


FIG. 6.

To illustrate the possible types of errors, let us notice that the graph in Fig. 7 is not reducible, because what appears as a C part with external photon lines ρ and μ is not in fact a C part comprised in the functions $C_{\mu\rho}$ so that this graph cannot be treated as reducible. When selecting the irreducible graphs for $\Phi_{\mu\nu}$ from the totality of graphs obtained, the following consideration is helpful. We have shown in Sec. 2, that $\theta_{\mu\rho}(p, p', 0)$

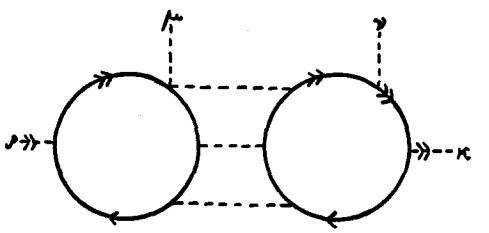


FIG. 7.

$=\theta_{\mu\rho}{}^b(p, p', 0)$, where $\theta_{\mu\rho}{}^b$ denotes the sum of the integrals corresponding to C parts with the photon line μ (with energy-momentum zero) necessarily belonging to the base line. Thus, if in a set of reducible $\Phi_{\mu\nu}$, only base-line C parts $\theta_{\mu\rho}{}^b$ appear, the irreducible skeleton for such $\Phi_{\mu\nu}$ contains the factor $\delta_{\mu\rho}$ and to obtain $\Phi_{\mu\nu}$ this factor can be replaced by $C_{\mu\rho}$ (the function for all and not merely base-line C parts) without incurring any error. The foregoing choice of "complementary" convention for p was designed, such that if a portion of a closed meson loop acts as the base line for graphs $\theta_{\mu\rho}$ (say) contained in a set of graphs $\Phi_{\mu\nu}$, the convention should insure (by differentiating the entire loop) that all relevant base line $\theta_{\mu\rho}{}^b$ do appear. The entire set $\theta_{\mu\rho}{}^b$ can then be replaced by a 4-vertex $\delta_{\mu\rho}$ and a (unique) skeleton thus defined.[18]

By categorizing the relevant graphs and using our inductive procedure, we shall establish the following equations (which are completely analogous to the set

[17] The choice of a convention is arbitrary but it is absolutely essential that once a convention is adopted it must be consistently followed.

[18] In this section (and subsequent sections) the distinction of δ_{ij} and δ_{ji} will be understood and the arguments will proceed in terms of 4-vertices, invoking the "weight factors" if necessary.

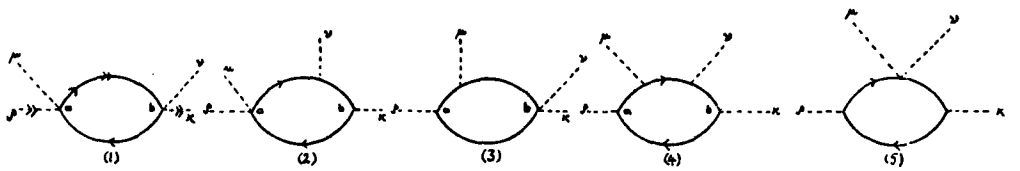


FIG. 8.

of Eqs. (4)):

$$\left.\begin{aligned} X_{\mu\nu}&=Z_3^{-1}X_{\mu\nu 1}(e_1)\\ W_\mu&=Z_3^{-1}W_{\mu 1}(e_1)\\ D_F'&=Z_3D_{F1}'(e_1).\end{aligned}\right\} \tag{31}$$

Except for the graphs in Fig. 8, it is possible to make unambiguous insertions (4) and (31) in all the lines and vertices of the irreducible $\Phi_{\mu\nu}$ (the vertices now including those in Fig. 5), so as to obtain all derived graphs and their corresponding integrals.

For the graphs in Fig. 8, however, insertion of certain types of C parts causes overlaps. Barring these, and considering all other irreducible graphs for $\Phi_{\mu\nu}$, we can express the sum of their integrals and of the reducible graphs derived from them in the fundamental form.

In Appendix I we will show, by a procedure of categorization, that the sum of the integrals from the graphs in Fig. 8 and all graphs derived from them can also be given in the fundamental form. Adding together the contribution from these and $2i\bar{\delta}_{\mu\nu}$ for the 4-vertex in Fig. 5(b), we obtain

$$X_{\mu\nu}=2i\bar{\delta}_{\mu\nu}+Z_3^{-1}[2iC(e_1)\bar{\delta}_{\mu\nu}+\Phi_{\mu\nu c}(e_1)].$$

At this stage a choice of $Z_3=1-C(e_1)$ establishes the relation, $X_{\mu\nu}=Z_3^{-1}X_{\mu\nu 1}(e_1)$. From this and (29) and (30) we prove as in Sec. 2 that $D_F'=Z_3D_{F1}'(e_1)$.

5. M PARTS

In this section the problem of M parts is dealt with. We show that with a correct choice of $\delta\lambda$ in a term $\delta\lambda\phi^{*2}\phi^2$ in the Hamiltonian, we cannot only consistently cancel all M divergences in the theory (whether they arise inside the graphs representing other primitive divergents or from what would otherwise be nondivergent graphs), we can also arrange that the graphs with 4-M vertices introduced into the theory by this term introduce no new infinities. It is easy to verify that if the $\delta\lambda$ term constituted the only interaction term the number of primitive divergents in the theory would be precisely two; namely, the graphs with two external meson lines (meson self-energy graphs) and those with four external meson lines (M parts).

The choice of $\delta\lambda$ is made in three steps.

(1) Consider all irreducible simple M parts. If M_d represents the sum of their true divergences, a choice of $\delta\lambda$ such that $\delta\lambda+M_d=0$ cancels all divergences arising from these graphs.

(2) For the case of irreducible nonsimple M parts, unlike scalar meson-nucleon theories, joining two simple M parts may lead to an $M-M$ overlap. An example is shown in Fig. 9. However, whether or not these overlaps occur, our subtraction procedure gives the result that the correct choice of $\delta\lambda$ is once again given by $\delta\lambda+M_d=0$, where $\delta\lambda$ is the sum of true divergences arising from all irreducible M parts, whether simple or otherwise. In spite of the overlap, the manner of proof is exactly similar to that given in detail in I, Sec. A, and is not repeated.

(3) To obtain all the original M parts from irreducible M parts, we make the usual insertions in all the lines and the vertices. Assuming (a) that the relations (3) hold and (b) that an insertion does not cause a further overlap, we immediately see by counting up the number of lines and vertices (including M vertices) that the correct final value for $\delta\lambda$ is given by $Z_2^{-2}M_d(e_1)+\delta\lambda=0$ where $M_d(e_1)$ is the sum of the true divergences from all M parts. Condition (b) is satisfied for all except the three irreducible graphs in Fig. 10.

In these graphs, insertion of certain types of C parts causes $C-M$ overlaps. By a procedure of categorization it is not difficult to prove that the contribution made by such graphs to $\delta\lambda$ is also found as $-Z_2^{-2}M_d$, where M_d is the sum of their true divergences.

To prove (a) we remark that the foregoing proof for the cancellation of M-divergences extends to the case when M parts are contained inside other graphs. In particular for photon self-energy graphs or C parts there exist other graphs in the theory derived by replacing the contained M parts by M vertices in all possible ways. These graphs combine to cancel the M divergences times the corresponding "reduced integrals" so that neither need be considered at any stage of our inductive procedure for the formation of the functions Π^* or $C_{\mu\nu}$.

Since our procedure for deriving relations for operators $\sum^*$ and Λ_μ is not to follow the relevant graphs but to obtain these analytically from $\theta_{\mu\nu}$ we have to show that the explicit neglect of M divergences and of the corresponding graphs with 4-M vertices in $\theta_{\mu\nu}$ does not affect the proof of the identities (16) and (17); in

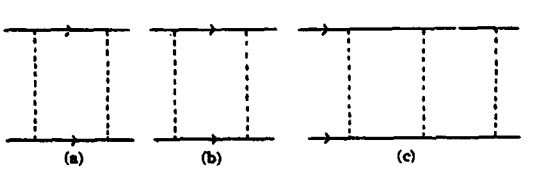


FIG. 9.

other words that this neglect is justified for both sides of the equations represented in (16) and (17).

The proof would be trivial if, for example, the differentiation $\partial/\partial p$ which, from the graphs corresponding to $\sum^*(p)$, produces all graphs relevant to $\Lambda_\mu(p, p)$ (in the manner of the proof in Sec. 3), also took all compensatory self-energy graphs with M vertices to corresponding compensatory vertex graphs. Since the proof of equality of $-(1/2\pi i)(\partial/\partial p)\sum^*$ with $\Lambda_\mu(p, p)$, for example, must depend on letting p run along the base line we have to show that a choice of basic momentum vectors which allows the path of p to lie along the base line also allows us to isolate M divergences.

The entire problem is closely linked with the problem of the general possibility of a "correct" choice of basic variables. This was discussed in detail in a previous paper[19] and only the results will be given here. For vertex parts situations arise when the number of divergences exceeds the number of subintegrations. In such cases, however, the extra divergences are M divergences which prove to be "final," and need not be separated. This happens, however, only on account of the gauge-invariance of the theory.

For meson self-energy graphs there is (as in the meson-nucleon case) another class of M divergences

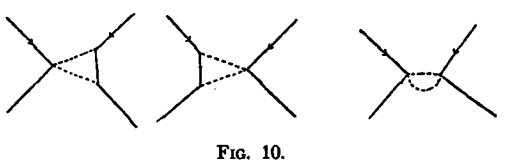


FIG. 10.

which were also called "final"[20] in Sec. B, I. The reduced integral corresponding to such "final" divergences is, by definition, independent of the external momentum p. Their effect is to leave just one type of graph with a 4-M vertex (illustrated in Fig. 6, I) and those derived from it by insertions in the meson line, as "odd" graphs which do not act as compensatory graphs like all other graphs with 4-M vertices, and so have to be considered separately. Here, as in the meson-nucleon case, these graphs only contribute to the mass renormalization constant, making it an explicit function of $\delta\lambda$; a fact finding analytic expression in that a differentiation $\partial/\partial p$ for such graphs gives the result zero.

Similar considerations apply for Π^*, Δ_μ, and $\Phi_{\mu\nu}$. The proof of the finiteness of the S matrix now follows precisely as in D II Sec. VII.

6. CONCLUSION

The only new feature arising for the renormalization of spin zero Bose particles interacting with the electromagnetic field is the introduction into the Hamiltonian of the (infinite) direct interaction term $\delta\lambda\phi^{*2}\phi^2$. In a sense it is satisfactory that this same term can remove the M divergences associated with the scattering of a meson by a meson when these particles interact through an exchange of virtual nucleons. The definition adopted in Sec. V of the divergent part of an M integral, however, makes $M_c(p_0, p_0, p_0)=0$, so that no graph of any order higher than the second in e^2, can contribute to the scattering amplitude of two mesons of equal initial and final momenta. Theoretically it is possible to proceed slightly differently and to introduce into the Hamiltonian, besides the compensatory $\delta\lambda\phi^{*2}\phi^2$, another "real" direct-interaction term $\lambda\phi^{*2}\phi^2$. The graphs involving 4-vertices with $\delta\lambda$ as coefficient are differentiated from those with λ as coefficient, the definition of original graphs being extended to include, besides the graphs involving 3-e- and 4-e^2-vertices, also the graphs containing 4-λ-vertices. The renormalization of the theory (not presented here in detail) proceeds as before, except that the Z factors now appear as functions of both e_1 and λ_1 (the renormalized value of λ) while the graphs with $\delta\lambda$-vertices compensate the additional M divergences introduced by the graphs containing these new λ-vertices as well. Thus $\delta\lambda$ is to be chosen as $\delta\lambda+Z_2^{-2}M_d(e_1, \lambda_1)=0$ while $\lambda_1=Z_2^{-2}\lambda$. The entire theory, after renormalization, appears in terms of two constants λ_1 and e_1. The retention of the condition $M_c(p_0, p_0, p_0)=0$ has the desirable feature, however, that even with this new term, the additional contribution to the Møller scattering amplitude, for mesons of equal initial and final momenta, is not given as a power series in λ_1 but consists merely of λ_1 itself.

The scheme presented here sketches what is theoretically possible. On account of the new feature noted, the physical validity of our renormalization scheme cannot be extrapolated from the fact that a very close agreement with experiment exists for the renormalized theory of spinor electrodynamics. To determine whether a constant λ_1 exists (even if only to find that its value is zero) and whether the answers to any physical problems given by the foregoing scheme approximate to the truth, we must turn again the pages of nature.

ACKNOWLEDGMENTS

The investigation was suggested by Dr. P. T. Matthews to whom the author is indebted for numerous discussions and detailed criticism of the manuscript. It gives the author great pleasure to express his profound indebtedness to Dr. N. Kemmer and Dr. F. J. Dyson for continual encouragement received throughout the progress of the work and to the Education Department of the Government of Punjab, Pakistan, and the Institute for Advanced Study, Princeton, for financial support.

[19] Abdus Salam, Phys. Rev. 83, 426 (1951).

[20] Opening any one photon line in a meson self-energy graph (or a meson line in a photon self-energy graphs) gives a C part with a corresponding "final" C divergence. Their effect is precisely similar to that of "final" M divergences.

APPENDIX I

Here we consider the graphs in Fig. 8 and all reducible graphs which can be derived from them and prove that their contribution to $\Phi_{\mu\nu}$ can be expressed in the fundamental form.

In the 4-vertices appearing in the graphs in Fig. 8 an insertion of those C parts which have their two external photon lines belonging either to the same 4-vertex (class B) or to two consecutive 3-vertices on the same meson line (class C), causes an overlap. These same C parts inserted at one end 4-vertex of the graphs in Fig. 8, appear simultaneously as an insertion at the other end. The situation is completely analogous to the insertion of vertex parts in end-vertices of photon (or electron) self-energy graphs in spinor electrodynamics (the "b divergences" of D II Sec. VII). Let the class of C parts $B+C$ be called A, while let T denote all C parts other than those in class A. From Sec. 2,

$$T_{\mu\nu}=Z_4^{-1}[T_d\delta_{\mu\nu}+T_{\mu\nu c}(e_1)].$$

We now categorize the graphs derived from those in Fig. 8. The categories obtained are similar to those in Sec. IV, (I). In graphs (1), (2), (3), a and b stand for the end 4-vertices. Graph (1) has both its end-vertices as 4-vertices, (2) and (3) have either the left- or the right-hand vertex (as drawn here) as a 4-vertex while (4) has no 4-vertex at all. Let the graphs (1), (2), (3), and (4) in Fig. 8 belong to category [1]. Insert all *irreducible* A at the 4-vertex $a_{[1]}$ in [1|1] and [1|2], thereby obtaining a set of graphs belonging to category [2]. These graphs can once again be distinguished as [2|1], [2|2], [2|3], and [2|4], according as their end-vertices are 3-vertices or 4-vertices. Insertion of irreducible A at the 4-vertices $a_{[2]}$ in [2|1] and [2|2] then gives all graphs in category [3] and so on. $a_{[n]}$ and $b_{[n]}$ stand for the end 4-vertices appearing in $[n|1]$, $[n|2]$, and $[n|3]$ at every stage.

Given a graph in $[n]$; let $A_a{}^1$ denote the true divergent constant arising from the irreducible A which was inserted at $a_{[n-1]}$ to obtain this particular graph in $[n]$; $A_a{}^2$, the constant of true divergence from the reducible graph A which could be inserted at $a_{[n-2]}$ to obtain this graph in $[n]$. This last graph is reducible because it contains as an insertion the irreducible graph with true divergence $A_a{}^1$. Similar definitions apply for $A_b{}^1$, $A_b{}^2$, $A_b{}^3\cdots$, etc. These definitions are analogous to those of $L_a{}^1$, $L_a{}^2$, $\cdots$, $L_b{}^1$, $L_b{}^2$, $\cdots$ in Sec. IV, I.

With these definitions we have the following lemma.

Lemma 1

$$\begin{aligned}[n]&=[n|1]+[n|2]+[n|3]+[n|4]\\&=A_a{}^1[n-1|1]+A_a{}^2[n-2|1]+A_a{}^3[n-3|1]+\cdots\\&\quad+A_b{}^1[n-1|1]+A_b{}^2[n-2|1]+A_b{}^3[n-3|1]+\cdots\\&\quad-[A_a{}^1A_b{}^1[n-2|1]+A_a{}^1A_b{}^2[n-3|1]+\cdots\\&\quad+A_a{}^2A_b{}^1[n-3|1]+\cdots]+A_a{}^1[n-1|2]+A_a{}^2[n-2|2]\\&\quad+\cdots+A_b{}^1[n-1|3]+A_b{}^2[n-2|3]+\cdots\\&\qquad+2iF_d\delta_{\mu\nu}+F_{\mu\nu c}.\end{aligned}\tag{1A}$$

$2i\delta_{\mu\nu}F_d$ gives the true divergence of $[n]$, while $F_{\mu\nu c}$ is the convergent part. The proof of the lemma is given exactly on the lines of the proof of Lemma 1 (Sec. IV, I) and is not repeated.

In any graph $[n]$, we can insert self-energy and vertex parts without causing any overlaps, while at all 4-vertices except $a_{[n]}$ and $b_{[n]}$, all C parts can be inserted unambiguously. Correspondingly, we replace Δ_F by $\Delta_F'=Z_2\Delta_{F1}'$, D_F by Z_2D_{F1}', $(p+p')_\mu$ by $Z_1^{-1}\Gamma_{\mu 1}$ in all the lines and 3-vertices of the graphs in these categories; while for all 4-vertices except $a_{[n]}$ and $b_{[n]}$, replace $\delta_{\mu\nu}$ by $Z_4^{-1}C_{\mu\nu 1}$. For $a_{[n]}$ and $b_{[n]}$, however, only the replacement of $\delta_{\mu\nu}$ by $\delta_{\mu\nu}+T_{\mu\nu}=Z_4^{-1}[(Z_4+T_d-1)\delta_{\mu\nu}+(\delta_{\mu\nu}+T_{\mu\nu c})]$ is analytically valid.

In terms of graphs, we obtain unambiguously by the insertions corresponding to these substitutions all graphs derived from the irreducible graphs in Fig. 8, no graph appearing more than once.

Let $\Sigma[n]^*$ represent the totality of graphs after these replacements, $[n]^\times$ represent graphs $[n]$ with $\Delta_{F1}'(e_1)$, $D_{F1}'(e_1)$, $\Gamma_{\mu 1}(e_1)$, and either $C_{\mu\nu 1}(e_1)$ or $\delta_{\mu\nu}+T_{\mu\nu c}(e_1)$ appearing for their lines and vertices, $[n]_a{}^\times$ represent all graphs with the above factors $\Delta_{F1}'(e_1)$, $D_{F1}'(e_1)$, etc., except at the 4-vertex a where factor $\delta_{\mu\nu}$ remains unchanged, and let similar definitions apply for $[n]_b{}^\times$ and $[n]_{ab}{}^\times$. Then precisely as in Lemma 2, the entire class of graphs $\Sigma[n]^*$ can be expressed compactly as

$$\begin{aligned}\Sigma[n]^*&=Z_3^{-1}\Sigma\{[n]^\times+(Z_4+T_d-1)([n]_a{}^\times+[n]_b{}^\times\\&\qquad+(Z_4+T_d-1)[n]_{ab}{}^\times)\}.\end{aligned}\tag{2A}$$

From Lemma 1, however, by summing up

$$\begin{aligned}\Sigma[n]^\times&=A_d{}^\times\{\Sigma([n]_a{}^\times+[n]_b{}^\times-A_d[n]_{ab}{}^\times)\}\\&\qquad+2iF_d{}^\times(e_1)+F_{\mu\nu c}{}^\times,\end{aligned}\tag{3A}$$

where $A_d{}^\times=A_a{}^{1\times}+A_a{}^{2\times}+A_a{}^{3\times}+\cdots=A_b{}^{1\times}+A_b{}^{2\times}+A_b{}^{3\times}+\cdots$. Here $A_a{}^{1\times}$, etc., denote the true divergences with Δ_{F1}', D_{F1}', etc., replacing Δ_F, D_F, etc. $F_d{}^\times(e_1)$ is the sum of the true divergences of $[n]^\times$.

We have shown in Sec. 2 that $Z_4=1-T_d-A_d{}^\times$. So substituting (3)A in (2)A we obtain

$$\Sigma[n]^*=Z_3^{-1}\{2iF_d{}^\times(e_1)\delta_{\mu\nu}+F_{\mu\nu c}{}^\times(e_1)\}.\tag{4A}$$

It still remains to consider the graph (5) in Fig. 8. The set consisting of graphs (1), (2), (3), and (4) was completely symmetrical regarding the classes B and C ($B+C=A$) but the graph (5) contains only one 4-vertex $\delta_{\mu\nu}$. This vertex will be called $e_{[1]}$. Inserting irreducible B at $e_{[1]}$ we obtain graphs belonging to [2|1] while insertion of irreducible C at $e_{[1]}$ gives [2|2]. These (irreducible B and C) can be inserted again at $e_{[2]}$ in [2|1] to obtain [3]=[3|1]+[3|2] and so on. The graph (5) in Fig. 8, itself constitutes the category [1|1], there being no graph in [1|2]. When the divergence in such a graph $[n]$ is being isolated, the "reduced" graph left after the true divergence corresponding to the C part (which C part necessarily belongs to class C) has been removed, will be denoted by $[n|3]$. A graph $[n|3]$ does not itself belong to the class of graphs $[n]$, because for $[n|3]$, $\rho\kappa$ must be a 4-vertex. This new 4-vertex will be called f. Similarly let $[n|4]$ denote the graph obtained by a simultaneous reduction of $[n]$ at both ends, so that in the graph $[n|4]$ both e and f appear as 4-vertices. Again a graph $[n|4]$ does not itself belong to the class of graphs $[n]$.

To obtain all graphs that can be derived from the graph (5), we construct the above categories $[n]$ and then in their lines and vertices make the replacements $\Delta_F=Z_2\Delta_{F1}'(e_1)$, etc., except at the 4-vertex e where $\delta_{\mu\nu}$ is replaced by $Z_4^{-1}[(Z_4+T_d-1)\delta_{\mu\nu}+(\delta_{\mu\nu}+T_{\mu\nu c}(e_1)]$. If $[n]^*$, $[n]^\times$ are defined as those $[n]$ in which, instead of Δ_F, for example, appear, respectively, the factors Δ_F' and Δ_{F1}', while $[n]_e{}^\times$ are $[n|1]^\times$ except that at the 4-vertex e, $\delta_{\mu\nu}$ is left completely unchanged, then by counting lines and vertices we obtain, as in (2)A

$$\Sigma[n]^*=Z_3^{-1}\{\Sigma([n]^\times+(Z_4+T_d-1)[n]_e{}^\times)\}.\tag{5A}$$

Now considering the implications of the subtraction procedure the result corresponding to Lemma 1 for $[n]$ is as follows.

Lemma 2

$$\begin{aligned}[n]&=[n|1]+[n|2]=A^1[n-1|1]+A^2[n-2|1]+A^3[n-3|1]\\&\quad+\cdots+C^1[n-1|3]+C^2[n-2|3]+C^3[n-3|3]\\&\quad+\cdots-(A^1C^1[n-2|4]+A^1C^2[n-3|4]+\cdots\\&\qquad+A^2C^1[n-3|4]+\cdots)+2iG_d\delta_{\mu\nu}+G_{\mu\nu c}.\end{aligned}\tag{6A}$$

By replacing Δ_F, by Δ_{F1}', etc., in $[n]$ and summing up, we obtain from the foregoing

$$\begin{aligned}\Sigma[n]^\times&=\Sigma\{A_d{}^\times[n]_e{}^\times+C_d{}^\times([n]_f{}^\times-A_d{}^\times[n]_{ef}{}^\times)\}\\&\qquad+2iG_d{}^\times(e_1)\delta_{\mu\nu}+G_{\mu\nu c}{}^\times(e_1).\end{aligned}\tag{7A}$$

$[n]_f{}^\times$ are $[n|3]$ with Δ_{F1}', etc., replacing Δ_F except that the 4-vertex f is completely unchanged, and $[n]_{ef}{}^\times$ are similarly obtained from $[n|4]$ except that for both e and f, the factors $\delta_{\rho\kappa}$ and $\delta_{\mu\nu}$ are not changed. Substituting for $\Sigma[n]^\times$ in Eq. (5)A,

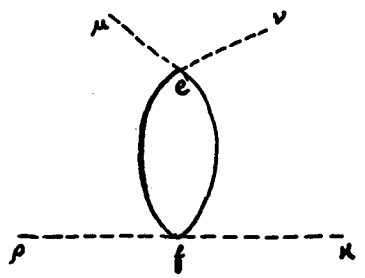


FIG. 11.

and using $Z_4=1-T_d-A_d^{\times}$ we obtain

$$\Sigma[n]^*=Z_3^{-1}(2i\delta_{\mu\nu}G_d^{\times}(e_1)+G_{\mu\nu e}^{\times}(e_1)) +Z_3^{-1}(\Sigma C_d^{\times}\{[n]_f^{\times}-A_d^{\times}[n]_{ef}^{\times}\}). \quad (8)A$$

We now prove the result:

Lemma 3

$$Z_3^{-1}(\Sigma C_d^{\times}\{[n]_f^{\times}-A_d^{\times}[n]_{ef}^{\times}\})=0. \quad (9)A$$

For the proof consider the graph in Fig. 11. Construct various categories from it, by inserting irreducible A, successively, at the vertex e only. The characteristic of these graphs then is that f is necessarily a 4-vertex for all of them. After these categories are constructed, make substitutions (4) in all lines and vertices, except in vertex e where $\delta_{\mu\nu}$ is replaced by $Z_4^{-1}[(Z_4+T_d-1)\delta_{\mu\nu}+(\delta_{\mu\nu}+T_{\mu\nu e})]$ and the 4-vertex f which is left completely unchanged. The result is that we obtain as the sum of integrals corresponding to all these graphs, precisely the expression,

$$Z_4Z_3^{-1}\Sigma([n]_f^{\times}+(Z_4+T_d-1)[n]_{ef}^{\times})(e_1),$$

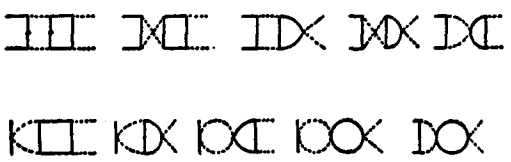


FIG. 12.

but this $=(\partial^2/\partial p_\mu\partial p_\nu)\int F(p+i, t^i, e)dtdt^i$ which is identically equal to zero, as can be seen by shifting $p+t\to t'$. Since $Z_4=1-T_d-A_d^{\times}\neq 0$ the proof is complete.

On account of Lemma 3, (8)A now gives

$$\Sigma[n]^*=Z_3^{-1}[2i\delta_{\mu\nu}G_d^{\times}(e_1)+G_{\mu\nu e}(e_1)], \quad (10)A$$

which is once again in the fundamental form. The required result is thus established.

APPENDIX II

The graphs R mentioned in Sec. 2 are a set of chain graphs "derived" from the 10 irreducible graphs in Fig. 12. Since C parts have two external photon lines, a chain of meson loops joined to each other by two photon lines can cause $C-C$ and $C-M$ overlaps. The proof that all possible types of $C-C$ overlaps are comprised in the classes P, Q, R is not difficult to give from very simple topological considerations. To show, further, that the sum of integrals from graphs R can be expressed in the fundamental form, a careful scheme of categorization needs to be developed. The general principles on which the proof proceeds is already illustrated in Sec. 2. The details are much more complicated but, as no new principles are involved, we shall not reproduce them here (Fig. 12).

Reprinted from THE PHYSICAL REVIEW, Vol. 90, No. 4, 690–695, May 15, 1953

Fredholm Theory of Scattering in a Given Time-Dependent Field

ABDUS SALAM, *St. John's College, Cambridge, England, and Government College, Lahore, Pakistan*

AND

P. T. MATTHEWS,* *Cavendish Laboratory, Cambridge, England*

(Received October 27, 1952)

It is shown that Feynman's relativistic solution for the scattering of an electron (or pair creation) by a given external field is the Fredholm resolvent of the related integral equation and is thus the unique and absolutely convergent solution for any strength of field.

INTRODUCTION

THE Fredholm theory of integral equations has been applied to the nonrelativistic theory of scattering by Jost and Pais.[1] We here consider the extension of this theory to the interaction of the quantized electron-positron field with a prescribed external electromagnetic field. This problem has been considered by Feynman.[2] Feynman's solution is most simply derived from the S matrix in the form given by Dyson.[3] The appropriate matrix element for electron scattering or pair creation is obtained as an expansion in the external field and is normalized by multiplying by the vacuum expectation value of the S matrix. We show that this is identical with the Fredholm resolvent of a related integral equation and is thus absolutely convergent for any strength of the external field, for which the cross section has any meaning.

In the first section the Fredholm theory is stated in a form given by Plemelj,[4] which exhibits the Fredholm solution in terms of the iterations of the kernel and its traces. These quantities have the advantage over the usual form of the theory[5] that they are either the same as, or closely related to, expressions occurring in the S matrix and can be written down directly by Feynman's graphical methods. The relation of the Fredholm solution to the solution by iteration is discussed. The problem of scattering in a pure external field is then treated in Secs. 2 and 3, with the result stated above. The case of a static field is related to the work of Jost and Pais.[1]

1. FREDHOLM THEORY

Consider Fredholm's integral equation

$$x(s)=y(s)+\lambda\int K(s,t)x(t)dt, \qquad (1.1)$$

$$(\text{or } x=y+\lambda Kx),$$

where the integration may be over a fixed interval, finite or infinite. Smithies[6] has shown that, if $K(s,t)$ is a measurable function of s and t, and

$$C_1\equiv\int\int|K(s,t)|^2dsdt<\infty, \qquad (1.2)$$

then (1.1) has the unique solution

$$\begin{aligned}x(s)&=d^{-1}(\lambda)\int D(\lambda,s,t)y(t)dt,\\&=d^{-1}(\lambda)\Delta(\lambda,s),\end{aligned} \qquad (1.3)$$

for all λ for which $d(\lambda)\neq 0$. Here

$$d(\lambda)=\sum_{n=0}^{\infty}d_n\lambda^n, \qquad (1.4)$$

$$D(\lambda,s,t)=\sum_{n=0}^{\infty}D_n(s,t)\lambda^n \qquad (1.5)$$

[$D(\lambda,s,t)$ is called the Fredholm resolvent], where $d_0=1$,

$$d_n=\frac{(-1)^n}{n!}\begin{vmatrix}\sigma_1 & n-1 & 0 & \cdots & 0\\ \sigma_2 & \sigma_1 & n-2 & \cdots & 0\\ \cdot & & & & \\ \cdot & & & & \\ \sigma_n & \sigma_{n-1} & & \cdots & \sigma_1\end{vmatrix}, \qquad (1.6)$$

$$D_n(s,t)=\frac{(-1)^n}{n!}\begin{vmatrix}\delta(s-t) & n & 0 & 0 & \cdots & 0\\ K(s,t) & \sigma_1 & n-1 & 0 & \cdots & 0\\ K^2(s,t) & \sigma_2 & \sigma_1 & n-2 & \cdots & 0\\ \cdot & & & & & \\ \cdot & & & & & \\ K^n(s,t) & \sigma_n & \sigma_{n-1} & & \cdots & \sigma_1\end{vmatrix} \qquad (1.7)$$

$$\begin{aligned}K^n(s,t)&=\int K^{n-1}(s,u)K(u,t)du\\&=\int K(s,u)K^{n-1}(u,t)du,\end{aligned} \qquad (1.8)$$

* Now at Department of Mathematical Physics, University of Birmingham, Birmingham, England.

[1] R. Jost and A. Pais, Phys. Rev. 82, 840 (1951).

[2] R. P. Feynman, Phys. Rev. 76, 749 (1949).

[3] F. J. Dyson, Phys. Rev. 75, 486, 1736 (1949).

[4] J. Plemelj, Monatsch. Math. 15, 93 (1904).

[5] See, for example, E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, 1940), fourth edition, Chapter XI.

[6] F. Smithies, Duke Math. J. 8, 107 (1941).

and

$$\sigma_n = \mathrm{Trace} K^n = \int K^n(s,s)ds. \tag{1.9}$$

If $x(s)$ is a spinor function, the trace in (1.9) is also taken over the spinor suffixes of K. The function $d(\lambda)$ is proved to be convergent for all complex values of λ, so that it is an integral function of λ; the resolvent, $D(\lambda, s, t)$, is also uniformly convergent for all λ, so the solution (1.3) is valid for the whole complex λ-plane (except for λ satisfying $d(\lambda)=0$). This formulation is shown to be identical with the usual formulation by Smithies.[6]

As already stated in the introduction, an advantage of this form is that for any field theoretic equations, the expressions $K^n y$ are just those appearing in the solution by iteration[7] and can be calculated directly by the Feynman-Dyson[2,3] method. The σ_n are closely related to the vacuum graphs of the S matrix, and Feynman's methods are again immediately applicable.

From the definition (1.6),

$$d_n = -\frac{1}{n}\sum_{m=1}^{n-1}\sigma_m d_{n-m}, \tag{1.10}$$

so that

$$\sum_{n=0}^{\infty}(n+1)d_{n+1}\lambda^n = -\Big(\sum_{n=0}^{\infty} d_n\lambda^n\Big)\Big(\sum_{q=0}^{\infty}\sigma_{q+1}\lambda^q\Big). \tag{1.11}$$

Thus

$$d^{-1}\frac{\partial d}{\partial \lambda} = -\sum_{q=0}^{\infty}\sigma_{q+1}\lambda^q.$$

Hence

$$d(\lambda) = \exp\Big[-\sum_{n=1}^{\infty}\sigma_n\lambda^n/n\Big]. \tag{1.12}$$

From the definition (1.7),

$$D_n = \sum_{m=0}^{n} K^m d_{n-m} \quad (= Id_n + KD_{n-1}). \tag{1.13}$$

Hence

$$D(\lambda) = d(\lambda)(1+\lambda K+\lambda^2K^2+\cdots). \tag{1.14}$$

The relation between the Fredholm solution and the solution by iteration is brought out clearly by substituting (1.14) into (1.3). The Fredholm solution then reduces to

$$x = \frac{D(\lambda)y}{d(\lambda)} = \frac{d(\lambda)(1+\lambda K+\lambda^2K^2+\cdots)y}{d(\lambda)}, \tag{1.15}$$

which is just the solution by iteration if the $d(\lambda)$ term is canceled. The Fredholm solution is exhibited as an analytic continuation of the iterated solution; the function $d(\lambda)$ is an integral function such that its zeros cancel the poles of the solution by iteration,[8] $(1+\lambda K+\lambda^2K^2+\cdots)y$, thus making $D(\lambda)y$ also an integral function. Fredholm's work shows not only that such a function exists but provides a method for actually calculating it.

The proof of the convergence of the Fredholm solution is based on the inequality

$$|\sigma_n| \leqslant |C_1|^{n/2} \quad (n=2,3,\cdots). \tag{1.16}$$

The condition (1.2) does not require that σ_1 be finite, and in fact the solution is unaltered by replacing σ_1 by zero in (1.6) and (1.7), whatever its actual value. However, if σ_2 diverges the condition (1.2) cannot be satisfied.

It may happen that $C_1, \cdots, C_{m-1}$ are not bounded, but that C_m is, where, by definition

$$C_m \equiv \int\int |K^m(s,t)|^2 ds dt. \tag{1.17}$$

One can still obtain a convergent solution to the equation by first iterating m times. Thus, in operator notation,

$$\begin{aligned} x &= y+\lambda Kx \\ &= (1+\lambda K+\lambda^2K^2+\cdots+\lambda^{m-1}K^{m-1})y+\lambda^m K^m x. \end{aligned} \tag{1.18}$$

This equation has a solution, since

$$C_m < \infty,$$

namely,

$$\begin{aligned} x = \frac{D_m(\lambda)y}{d_m(\lambda^m)} &= \frac{d_m(\lambda^m)}{d_m(\lambda^m)}\cdot(1-\lambda^mK^m)^{-1} \\ &\quad\times(1+\lambda K+\cdots+\lambda^{m-1}K^{m-1})y \\ &= \frac{d_m(\lambda^m)}{d_m(\lambda^m)}\cdot(1+\lambda K+\lambda^2K^2+\cdots)y, \end{aligned} \tag{1.19}$$

where

$$d_m(\lambda^m) = \exp\Big\{-\Big[\lambda^m\sigma_m+\frac{\lambda^{2m}}{2}\sigma_{2m}+\cdots\Big]\Big\}. \tag{1.20}$$

[7] The solution by iteration is known as the Neumann-Liouville solution. This is the Born approximation in nonrelativistic scattering theory and is the "weak coupling" expansion of relativistic field theory.

[8] This result has been established quite rigorously in the literature. Just to clarify the structure of the Fredholm solution, we can cast $d(\lambda)$ in a different form by the following nonrigorous argument. This was pointed out to the authors by Professor R. E. Peierls.

The solution by iteration is the binomial expansion of

$$x = \frac{1}{1-\lambda K}y.$$

The poles of this solution are given by the eigenvalues of the homogeneous equation

$$(1-\lambda_\alpha K)x_\alpha = 0.$$

The eigenvalues of K^n are $(1/\lambda_\alpha)^n$, and the trace σ_n is thus

$$\sigma_n = \Sigma_\alpha\,(1/\lambda_\alpha)^n.$$

Hence, by Eq. (1.12),

$$\begin{aligned} d(\lambda) &= \exp\Big[-\sum_{n=1}^{\infty}\sum_{\alpha}\frac{1}{n}\Big(\frac{\lambda}{\lambda_\alpha}\Big)^n\Big] \\ &= \exp[\Sigma_\alpha \log(1-\lambda/\lambda_\alpha)] \\ &= \Pi_\alpha(1-\lambda/\lambda_\alpha), \end{aligned}$$

which is just the form required for $d(\lambda)$ to cancel the poles of the solution by iteration.

This solution is reducible, and one can factor an integral function from the numerator and the denominator. Poincare[9] has shown the irreducible solution is given by Eqs. (1.3)–(1.9) with $\sigma_1, \sigma_2, \cdots \sigma_{m-1}$ replaced by zero.

2. THE EXTERNAL FIELD

Consider the problem of the quantized Dirac field in interaction with a given external field, which is time dependent and may create pairs. The field produced by the electrons themselves is neglected. The S matrix for the system is

$$S=\sum_n \frac{(-i)^n}{n!}\int\cdots\int dx_1\cdots dx_n \times P(H^e(x_1),\cdots H^e(x_n)), \quad (2.1)$$

where

$$H^e(x)=-i\lambda\bar{\psi}(x)A^e(x)\psi(x), \quad (2.2)$$

$$A^e(x)=A^e{}_\mu(x)\gamma_\mu, \quad (2.3)$$

and λ is a constant which determines the strength of the field.

We consider the cross section for scattering of an electron from a state $u(q)$ to a state $u(p)$, where both momentum and spin are specified. The transition amplitude obtained from the S matrix is written as $(p|S|q)$. From Dyson's theory[9] applied to (2.1) it follows that

$$(p|S|q)=(0|S|0)(p|R|q); \quad (2.4)$$

where $(0|S|0)$ is the vacuum expectation value of the S matrix, represented by vacuum graphs, while $(p|R|q)$ is, by definition, represented by the connected graphs for the scattering process itself.[10] Because we are dealing with a time-dependent external field the vacuum-to-vacuum transition probability $|(0|S|0)|^2$ is not unity, and so $(0|S|0)$ is not a mere phase factor and has to be expressly considered.

To calculate $(p|R|q)$ define $M(p, q)$ such that

$$(p|R|q)=\bar{u}(p)M(p, q)u(q). \quad (2.5)$$

From the S matrix an integral equation for $M(p, q)$ can readily be derived. For weak external fields (λ small), where an expansion in field strength is permissible we have, by Feynman-Dyson methods,

$$M(p, q)=\lambda A^e(p-q)-2\pi\lambda^2\int A^e(p-k)S_F(k) \times A^e(k-q)d^4k, +\cdots, \quad (2.6)$$

where

$$S_F(k)=\frac{1}{2\pi i}\frac{i\gamma_\mu k_\mu-\kappa}{k^2+\kappa^2-i\epsilon}, \quad \epsilon>0. \quad (2.7)$$

Write $M_I(p, q)$ for the right-hand side of (2.6) and define

$$(p|R_I|q)=\bar{u}(p)M_I(p, q)u(q). \quad (2.8)$$

Now $M_I(p, q)$ is the solution by iteration (hence the suffix I) of the integral equation[11]

$$M(p, q)=\lambda A^e(p-q)-2\pi\lambda\int A^e(p-k)S_F(k) \times M(k, q)d^4k, \quad (2.9)$$

which defines $M(p, q)$ for any strength of field. This is an equation for $M(p, q)$, (q constant) of the Fredholm type and we can consider the application of the theory of the previous section.[12]

The kernel of the equation is

$$K(p, k)=A^e(p-k)S_F(k)(-2\pi). \quad (2.10)$$

This is a measurable function for reasonable $A^e(p-k)$ provided the limit $\epsilon\to 0$ is not taken until the end of the calculation.[13]

The Fredholm traces σ_n can be looked at graphically, as stated in the introduction. For this particular integral equation, the σ's happen to be precisely the expressions represented by the vacuum graphs of the S matrix theory. Define

$$L=\sum_{n=1}^{\infty}\frac{\sigma_{2n}\lambda^{2n}}{2n} \quad (2.11)$$

so that the Fredholm determinant [Eq. (1.12)] is

$$d(\lambda)=\exp(-L). \quad (2.12)$$

From (2.11), L is the sum of all vacuum graphs taken singly.[14] With this definition of L, Feynman has shown that

$$(0|S|0)=\exp(-L), \quad (2.13)$$

[9] H. Poincare, Acta Math. 33, 57 (1910).

[10] Dyson's treatment is a restatement in more conventional terms of Feynman's solution. $(p|S|q)$ is called the "absolute" probability amplitude and $(p|R|q)$ the "relative" probability amplitude by Feynman.

[11] This equation has already been considered by M. Neuman, Phys. Rev. 83, 1258 (1951); 85, 129 (1952).

[12] Since $S_F{}^A(x, y)$, the electron propagator in a given field, satisfies

$$\left[\gamma_\mu\frac{\partial}{\partial x_\mu}-\kappa-ieA(x)\right]S_F{}^A(x, y)=-2i\delta(x-y),$$

$$S_F{}^A(x, y)=S_F(x-y)-\frac{e}{2}\int S_F(x-z)A(z)S_F{}^A(z, y)dz.$$

This is the same equation and can be solved by the same methods.

[13] We refer here to the standard procedure for calculating Feynman integrals for σ_n and K^n, in terms of which the answer is expressed [formulas (1.3)–(1.9)]. This procedure has been demonstrated for particular examples by R. P. Feynman [Phys. Rev. 76, 769 (1949); see particularly the penultimate paragraph of Sec. 7] and discussed generally by R. J. Eden, Proc. Roy. Soc. (London) A210, 388 (1952).

For $\epsilon>0$ the momentum integrals are all taken along the real axis. After completing these integrals, it is, in general, necessary to keep $\epsilon>0$ in order to define the analytic continuation of these functions round the branch points which occur at creation thresholds. Only after this can the limit $\epsilon\to 0$ be taken, to give the physical answer.

[14] Note that $\sigma_{2n+1}=0$ by Furry's theorem. The factor $(1/2n)$ in the nth term comes from $(2n!)^{-1}$ in (2.1) multiplied by $(2n-1)!$, the number of permutations of the $2n$ points round the closed loop of the graph for σ_{2n}.

and, therefore,[15]

$$(0|S|0)=d(\lambda). \tag{2.14}$$

Let us now consider the σ's for convergence. The expression for σ_2 is

$$\sigma_2=(2\pi)^2\int\int dpdk \times \mathrm{Tr}[A^e(p-k)S_F(k)A^e(k-p)S_F(p)]. \tag{2.15}$$

By a change of variable,

$$\sigma_2=(2\pi)^2\int dt A_\mu^e(t)A_\nu^e(-t)\int dk \times \mathrm{Tr}[\gamma_\mu S_F(k)\gamma_\nu S_F(t-k)]. \tag{2.15'}$$

The integral over k is the expression which occurs in the vacuum polarization and is divergent. (This is obvious from the graph of σ_2, which is the well-known loop of two electron lines.) Thus, σ_2 diverges for any choice of A^e. By (1.16) it follows that

$$C_1>\infty, \tag{2.16}$$

and Fredholm theory is not immediately applicable to the equation in this form. This difficulty can be overcome by the device of Poincare,[8] stated at the end of the previous section. Fredholm theory can be applied in a slightly modified form to our equation, provided[16]

$$C_2<\infty. \tag{2.17}$$

For the moment we will assume this condition to be satisfied and will show in the next section how all fields of physical interest can be included.

Assuming (2.17), the unique convergent solution for any strength of field is given by replacing σ_2 by zero in the general formulas of Sec. 1. By (1.3), (2.8), and (2.12) this is

$$(p|R|q)=\{\bar{u}(p)\Delta_q(\lambda, p)u(q)\}/\exp(-L'), \tag{2.18}$$

where

$$L'=\sum_{n=2}^{\infty}[\sigma_{2n}\lambda^{2n}/2n], \tag{2.19}$$

and Δ is defined by (1.7) applied to (2.10). The suffix q on Δ_q denotes that it is also a function of q. By (2.11) and (2.13)

$$(p|S|q)=\exp(-L)\{\bar{u}(p)\Delta_q(\lambda, p)u(q)\}/\exp(-L') =\{\bar{u}(p)\Delta_q(\lambda, p)u(q)\}\exp[-\sigma_2\lambda^2/2]. \tag{2.20}$$

The factor in curly brackets is essentially the resolvent of the Fredholm solution of (2.9) and is thus an integral function of λ. The σ_2 which occurs in the other term is infinite, but its real part is finite,[17] and it thus contributes a finite numerical factor to the probability $|(p|S|q)|^2$. Our final expression for $|(p|S|q)|^2$ is thus an infinite series in λ, each term of which is finite, and which is absolutely and uniformly convergent for any value of λ, that is, for any strength of field.

But according to the form of $\Delta(\lambda)$ given by (1.15),

$$\Delta_q(\lambda, p)=\exp(-L')M_I(p, q). \tag{2.21}$$

Thus, the Fredholm solution, Eq. (2.20), can be written

$$(p|S|q)=(p|R_I|q)\exp(-L). \tag{2.22}$$

This is precisely the solution by iteration, properly normalized, which is derivable from (2.1) by standard methods and was originally given by Feynman.[2] This is a rather surprising result, since the iteration $(p|R_I|q)$ has, in general, a finite radius of convergence, and from Feynman's point of view the term $\exp(-L)$ is introduced simply as a normalizing factor. This identity of the normalized iterated solution and the Fredholm solution is a special property of the system under discussion and is a direct consequence of the relation (2.14) between the Fredholm determinant and the vacuum expectation value of the S matrix. The identity is not true for static external fields, which, for reasons which will be made clear in the next section, cannot be regarded as a particular case of the time dependent field.

We remark finally that the above theory can equally well be applied to the calculation of pair creation. Also, since electrons interact with each other explicitly only through the exclusion principle on the initial and final states, we can consider any number of electrons or positrons. This has been shown by Feynman.[2]

3. FREDHOLM CONDITION AND THE STATIC FIELD

A. The Fredholm Condition

We have yet to establish the condition (2.17), which justifies the application of the Fredholm theory to (2.9). From (1.17) and (2.10),

$$C_2=\int dk_1dk_2dldm \times|A(l-k_1)S_F(k_1)A(k_1-m)S_F(m) \times A(l-k_2)S_F(k_2)A(k_2-m)S_F(m)|. \tag{3.1}$$

[15] This result was obtained previously by M. Neuman. See reference 11.

[16] We are actually employing a generalization of Poincare's result, which has been proved by numerous authors [F. Smithies (private communication)].

[17] This has been calculated directly by J. Schwinger and others. The result in our notation is quoted by R. Karplus and N. M. Kroll, Phys. Rev. 77, 536 (1950). That the real part is finite can be inferred, however, from the unitarity of the S matrix. Let

$$S=1+eS_1+e^2S_2+\cdots.$$

Since S is unitary,

$$S_2+S_2^*=-S_1S_1^*.$$

Therefore,

$$\mathrm{Rl}[2\sigma_2]=(0|S_2+S_2^*|0)=-\sum_{p,q}|(0|S_1|p, q)|^2,$$

where the right-hand side is the total probability of creating one pair, calculated to lowest order. This is finite.

It must be shown that this is convergent. It is not true of C_2 as it was of C_1, that by a change of variables one can produce a divergent factor, independent of $A(k)$. Dyson has shown that the convergence of such integrals can be estimated by counting powers. On this basis it can be seen that C_2 converges, provided $A(k)$ falls off faster than k^{-3} for large k.

Consider the expression

$$E=\int J_\mu(k)A_\mu(k)d^4k=\int k^2A_\mu(k)A_\mu(-k)d^4k, \quad (3.2)$$

where J_μ is the current producing the external field. By a Fourier transformation this is equal to the energy density of the field, integrated over all space and time.

Now the transition probability

$$\omega_\infty=|(p|S|q)|^2 \quad (3.3)$$

is the probability for the whole of time and the whole of space. For purely general fields this is infinite and there is, in fact, no physically significant cross section that one can consider for such a field. We consider now the space extension only. In most cases the field will be defined throughout space, but will be effectively nonzero only over a finite region. In such a case the probability for the whole of space gives the probability for the whole field, which is the physically significant quantity. Alternatively the field may have some general periodicity in a lattice with some cell volume v. One may then ask for the cross section per cell. This may be calculated by considering the probability for all space for a field equal to the given field over a large but finite volume $V(\gg v)$ and zero elsewhere. The boundary should be made smooth and taken large enough for surface effects to be negligible.

The treatment of the time is exactly analogous. Either the field is defined for all time in such a way that it is effectively nonzero only for a finite period (in this case ω_∞ is, directly, the physically significant expression) or it has a periodicity τ. In this latter case the probability per period τ can be calculated from ω_∞, for the field switched on for a long, but finite time $T(\gg\tau)$.

In any significant case, the field considered is nonzero only over a finite region and for a finite time. For such fields the energy integral E is finite. But this integral is only convergent if $A(k)$ falls off faster than k^{-2} for large k.[18a] Therefore, for any time-dependent field for which a physically reasonable cross section can be considered, we have

$$C_2<\infty, \quad (3.4)$$

and Fredholm theory is applicable.

B. The Static Field

For a static field the expression of physical significance is not ω_∞, but ω, the probability per unit time. To derive this exactly from the S matrix, Eq. (2.1), it is necessary to consider the field to be on for an infinite time. Since the argument above depended essentially on the fact that the field was only "on" for a finite time, the static field requires special consideration.

A static field is

$$A_\mu(k)=A_\mu(\mathbf{k})\delta(k_0), \quad (3.5)$$

where $\mathbf{k}$ is the space part of k. Energy is conserved in any real process, so that

$$|(0|S|0)|^2=1.$$

Thus, $(0|S|0)$ is simply an (infinite) phase factor and probability amplitudes are given directly by $(p|R|q)$. Define $M'(\mathbf{p},\mathbf{q})$ by

$$M(p,q)=M'(\mathbf{p},\mathbf{q})\delta(p_0-q_0). \quad (3.6)$$

The solution by iteration for $M'(\mathbf{p},\mathbf{q})$ and hence $(p|R_I|q)$ is obtained by substituting (3.5) and (3.6) into (2.6).[18] The integral equation for $M'(\mathbf{p},\mathbf{q})$ is, from (2.9),

$$M'(\mathbf{p},\mathbf{q})=\lambda A^*(\mathbf{p}-\mathbf{q})-2\pi\lambda^2\int A^*(\mathbf{p}-\mathbf{k})S_0(\mathbf{k})\times M'(\mathbf{k},\mathbf{q})d\mathbf{k}, \quad (3.7)$$

where

$$S_0(\mathbf{k})=S_F(\mathbf{k},p_0) \quad (3.8)$$

is given by (2.7) with k_0 replaced by p_0. The Fredholm solution of (3.7) can be performed analogously to the solution of (2.10). The condition for the applicability of Fredholm theory is the convergence of the three-dimensional analog of (3.1). This can be deduced from the convergence of

$$\int|\mathbf{k}|^2A_\mu(\mathbf{k})A_\mu(-\mathbf{k})d\mathbf{k}. \quad (3.9)$$

This is the integral of the energy density over all space (but not time), which must converge for any physically reasonable field, as in the previous section.

From (3.6),

$$(p|R|q)=\bar{u}(\mathbf{p})M'(\mathbf{p},\mathbf{q})u(\mathbf{q})\delta(p_0-q_0)=(\mathbf{p}|\mathbf{R}|\mathbf{q})\delta(p_0-q_0). \quad (3.10)$$

Thus,

$$\omega_\infty=|(\mathbf{p}|\mathbf{R}|\mathbf{q})|^2\delta(p_0-q_0)\delta(0). \quad (3.11)$$

This is, of course, infinite, but the final factor is

$$\delta(0)=\lim_{T\to\infty}\lim_{x\to0}\frac{1}{2\pi}\int_{-T/2}^{T/2}e^{ixt}dt=\lim_{T\to\infty}(T/2\pi), \quad (3.12)$$

where T is the time for which the field is on. The probability per unit time is

$$\omega=\lim_{T\to\infty}(\omega_\infty/T)=\frac{1}{2\pi}|(\mathbf{p}|\mathbf{R}|\mathbf{q})|^2\delta(p_0-q_0). \quad (3.13)$$

[18a] This is not true if $A(x)$ is singular at the origin, but such cases may be included by introducing a suitable factor to smooth out the singularity.

[18] R. H. Dalitz, Proc. Roy. Soc. (London) **A206**, 509 (1951).

Note that owing to the extraction of a δ-function factor in the derivation of (3.7) from (2.10), it is *not* true that the Fredholm determinant $d(\lambda)$ of (3.7) is equal to $(0|S|0)$. Consequently, for a static field, the iterated solution, $(\mathbf{p}|\mathbf{R}_I|\mathbf{q})$, (Born approximation) is *not* the same as the Fredholm solution. This is in agreement with the conclusion of Jost and Pais.[1]

The authors are indebted to Dr. N. Kemmer for numerous discussions and particularly Dr. F. Smithies for much help with the theory of integral equations. One of us (A.S.) would like to thank the Education Department of the Punjab Government, Pakistan for a grant enabling him to stay at St. John's College, Cambridge.

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 221, pp. 128–134, 1953

Covariant Fock equations

BY P. T. MATTHEWS
Department of Mathematical Physics, University of Birmingham

AND ABDUS SALAM
Government College, Lahore, Pakistan

(*Communicated by R. E. Peierls, F.R.S.—Received* 30 *July* 1953)

A new type of amplitude—Feynman amplitude—is defined for any state. These amplitudes are shown to propagate in the Feynman manner. It is shown further that they satisfy an infinite set of four-dimensional, covariant, coupled integral equations, analogous to the three-dimensional Fock equations between the ordinary probability amplitudes.

1. INTRODUCTION

The separation of the finite and physically significant parts of infinite expressions occurring in the interaction of coupled fields has so far been expressed in the framework of the S-matrix. This is, in general, applicable only to high-energy scattering problems, and for the meson-nucleon interaction it is not even applicable to these. Attempts to get away from charge expansions have been based on the Fock (1932) equations between the three-dimensional probability amplitudes of the state under consideration (Tamm 1945; Dancoff 1950). The infinities related to vacuum, self-energy and vertex effects are mixed up in these equations in an inextricably complicated manner. It is our object here to present an infinite set of coupled integral equations, which are a covariant analogue to the Fock equations, and which are obviously renormalizable in the restricted sense that no vacuum effects ever occur, and all other infinities appear directly in the S-matrix form, that is, as functions of Feynman propagators integrated over the whole of space-time. The finite parts can thus be defined directly. However, it still remains to be proved that the dropping of infinite terms can be done consistently, independent of a charge expansion. This point is not discussed here.

It should be mentioned that these equations relate a new type of amplitude—Feynman amplitude*—whose physical meaning is far from clear. This is the price one pays for an 'obviously renormalizable' form of the theory.

2. FEYNMAN AMPLITUDES

Consider the pseudo-scalar interaction of pseudo-scalar mesons with nucleons. Define the Dirac and Klein-Gordon operators

$$\mathscr{D}_x = \mathrm{i}\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + \kappa\right), \tag{1}$$

$$\overline{\mathscr{D}}_x = \mathrm{i}\left(-\gamma_\mu^T \frac{\partial}{\partial x_\mu} + \kappa\right), \tag{2}$$

$$\mathscr{K}_x = -\mathrm{i}(\Box_x - \mu^2); \tag{3}$$

* The name is due to Gell-Mann & Goldberger (1952) who discussed a restricted class of these amplitudes with a slightly different definition. We are deeply indebted to Dr Gell-Mann for an exposition of his work, which has never been published in detail.

then the field equations are

$$\left.\begin{aligned} \mathscr{D}\psi(x) &= g\gamma(x)\,\psi(x)\,\phi(x), \\ \overline{\mathscr{D}}\overline{\psi}(x) &= g\gamma(x)\,\overline{\psi}(x)\,\phi(x), \end{aligned}\right\} \tag{4}$$

and

$$\mathscr{K}\phi(x) = g\gamma(x)\,\overline{\psi}(x)\,\psi(x),$$

where $\gamma(x)$ denotes γ_5 which must be combined with the ψ field in the form $\gamma\psi(x)$, $\overline{\psi}(x)\gamma$ or $\overline{\psi}(x)\gamma\psi(x)$. For the bare vacuum expectation value of Wick's (1950) T-product of interaction representation variables $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ we have

$$\langle T(\boldsymbol{\psi}(x)\,\overline{\boldsymbol{\psi}}(y))\rangle_0 = -\tfrac{1}{2}S_F(x-y) = s(x-y), \tag{5}$$

and

$$\langle T(\boldsymbol{\phi}(x)\,\boldsymbol{\phi}(y))\rangle_0 = \tfrac{1}{2}D_F(x-y) \;\; = d(x-y). \tag{6}$$

The functions $s(x)$ and $d(x)$ are introduced simply to absorb numerical factors. We also define

$$\bar{s}(x) = s(-x). \tag{7}$$

Then

$$\left.\begin{aligned} \mathscr{D}(x)\,s(x) &= \delta(x), \\ \overline{\mathscr{D}}(x)\,\bar{s}(x) &= \delta(x), \end{aligned}\right\} \tag{8}$$

and

$$\mathscr{K}(x)\,d(x) = \delta(x).$$

Denote the normal (N-) product* of several interaction representation variables by

$$\mathbf{N}(\boldsymbol{\psi}(x_i)\ldots;\ \overline{\boldsymbol{\psi}}(y_j)\ldots;\ \boldsymbol{\phi}(z_k)\ldots) \equiv \mathbf{N}(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k\ldots), \tag{9}$$

and use a similar notation for a T-product. Then Wick (1950) has shown that

$$\begin{aligned} T(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k\ldots) = {} & \mathbf{N}(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k\ldots) \\ & + \sum_{i<j} s(x_i-y_j)\,\mathbf{N}(\mathbf{x}_i^{-1}\ldots;\ \mathbf{y}_j^{-1}\ldots;\ \mathbf{z}_k\ldots) \\ & + \sum_{k<l} d(z_k-z_l)\,\mathbf{N}(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k^{-1}\ldots\mathbf{z}_l^{-1}\ldots) \\ & + \ldots, \end{aligned} \tag{10}$$

where $\mathbf{N}(\mathbf{x}_i^{-1}\ldots)$ means that the factor $\boldsymbol{\psi}(x_i)$ should be dropped from the N-product. The remaining terms on the right-hand side of (10) are made up of all pairs of s or d factors multiplied by the N-product of the remaining operators, plus all triplets of s's and d's times the N-product of the remaining operators, plus, etc. This expression can be reversed to express an N-product in terms of T-products.

THEOREM 1:

$$\begin{aligned} \mathbf{N}(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k\ldots) = {} & T(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k\ldots) \\ & - \sum_{i,j} s(x_i-y_j)\,T(\mathbf{x}_i^{-1}\ldots;\ \mathbf{y}_j^{-1}\ldots;\ \mathbf{z}_k\ldots) \\ & - \sum_{k,l} d(z_k-z_l)\,T(\mathbf{x}_i\ldots;\ \mathbf{y}_j\ldots;\ \mathbf{z}_k^{-1},\mathbf{z}_l^{-1}\ldots) \\ & + \Sigma\, s(x_i-y_j)\,d(z_k-z_l)\,T(\mathbf{x}_i^{-1}\ldots;\ \mathbf{y}_j^{-1}\ldots;\ \mathbf{z}_k^{-1}\mathbf{z}_l^{-1}\ldots) \\ & + \ldots. \end{aligned}$$

This can be verified by substituting (10) back into the right-hand side.

* N-product is called S-product by Wick and denoted by colons. We use the nomenclature and notation of Dyson (1951).

So far we have been dealing with interaction representation variables. These can be split into annihilation and creation parts and the N-products are defined in terms of them. This is not so for Heisenberg operators. We *define* the normal product of Heisenberg operators $\psi(x_i)$, $\overline{\psi}(y_j)$ and $\phi(z_k)$ to be

$$\begin{aligned} N(x_i \ldots; y_j \ldots; z_k \ldots) \\ &= T(x_i \ldots; y_j \ldots; z_k \ldots) \\ &\quad - \Sigma s(x_i - y_j)\, T(x_i^{-1} \ldots; y_j^{-1} \ldots; z \ldots) \\ &\quad + \ldots, \end{aligned} \tag{11}$$

that is, by the right-hand side of theorem 1 with interaction operators replaced by Heisenberg operators in the T-products.

Let $\Psi\rangle$ be an eigenstate of the system and $\langle\Psi_0$ the true vacuum. Then define

$$A(x_1 \ldots x_n; y_1 \ldots y_l; z_1 \ldots z_m) = \langle\Psi_0 N(x_1 \ldots x_n; y_1 \ldots y_l; z_1 \ldots z_n)\, \Psi\rangle \tag{12}$$

to be the Feynman amplitude of the state $\Psi\rangle$. It is the main purpose of this paper to show that these amplitudes propagate in the Feynman (1949) manner, and, in fact, satisfy an infinite set of coupled integral equations, which are the four-dimensional analogue of the Fock equations. If all the times appearing in the equation are put equal to the origin of time at which the Heisenberg, interaction, and Schrödinger representations coincide, the Feynman amplitude reduces to a linear sum of the modified Fock amplitudes recently proposed by Dyson (1953). If, in addition, the true vacuum is replaced by the bare vacuum, one obtains the ordinary probability amplitudes of the Fock equations. The 'wave function' of the Bethe-Salpeter (1951) equation is a simple example of a Feynman amplitude (Gell-Mann & Low 1951).

3. The differential equations

In order to derive the differential equations satisfied by the Feynman amplitudes we state a second theorem.

Theorem 2:

(*a*) $$\begin{aligned} \mathscr{D}_i T(x_i \ldots; y_j \ldots; z_k \ldots) \\ &= g\gamma(x_i)\, T(x_i \ldots; y_j \ldots; x_i z_k \ldots) \\ &\quad + \sum_j \delta(x_i - y_j)\, T(x_i^{-1} \ldots; y_j^{-1} \ldots; z_k \ldots), \end{aligned}$$

(*b*) $$\begin{aligned} \overline{\mathscr{D}}_j T(x_i \ldots; y_j \ldots; z_k \ldots) \\ &= g\gamma(y_j)\, T(x_i \ldots; y_j \ldots; y_j z_k \ldots) \\ &\quad + \sum_i \delta(y_j - x_i)\, T(x_i^{-1} \ldots; y_j^{-1} \ldots; z_k \ldots), \end{aligned}$$

(*c*) $$\begin{aligned} \mathscr{K}_k T(x_i \ldots; y_j \ldots; z_k, z_l \ldots) \\ &= g\gamma(z_k)\, T(z_k, x_i \ldots; z_k, y_j \ldots; z_k^{-1}, z_l \ldots) \\ &\quad + \sum \delta(z_k - z_l)\, T(x_i \ldots; y_j \ldots; z_k^{-1}, z_l^{-1} \ldots). \end{aligned}$$

The detailed proof is given in appendix 1, but it is clear that the g-terms come from the equations (4) and the δ-functions from the derivatives of the ϵ-factors contained in the T-product.

The application of this theorem to the Feynman amplitude defined by (11) and (12) gives directly an expression for, say, $\mathscr{D}_i A(x_i \ldots; \ldots; \ldots)$ in terms of T-products. It is then possible to prove the important result that the right-hand side of such an equation is always exactly expressible in terms of other Feynman amplitudes. The proof follows directly from the definitions. It is clumsy to express in general and we will state here only the result for the D operator, namely,

$$\begin{aligned}
&\mathscr{D}_i A(x_i \ldots; y_j \ldots; z_k \ldots) \\
&\quad = g\gamma(x_i)\,[\sum_k d(z_k - x_i)\, A(x_i \ldots; y_j \ldots; z_k^{-1} \ldots) \\
&\qquad + \sum_j s(x_i - y_j)\, A(x_i^{-1} \ldots; y_j^{-1} \ldots; x_i, z_k \ldots) \\
&\qquad + A(x_i \ldots; y_j \ldots; x_i, z_k \ldots) \\
&\qquad + \sum_{j,k} d(z_k - x_i)\, s(y_j - x_i)\, A(x_i^{-1} \ldots; y_j^{-1} \ldots; z_k^{-1} \ldots)].
\end{aligned} \tag{13}$$

There are two similar equations for $\bar{\mathscr{D}}_j$ and $\mathscr{K}_k$. These will be referred to as *the* differential equations. These differential equations have a simple interpretation in terms of Feynman graphs. To see this it is useful to express each in an integral form by the following device. We have

$$\begin{aligned}
\mathscr{D}_x A(x, \ldots) &= F(x, \ldots) \\
&\equiv \int \delta(x - w)\, F(w, \ldots)\, \mathrm{d}w.
\end{aligned} \tag{14}$$

Operating by the inverse operator D_x^{-1} and using (8) gives

$$A(x, \ldots) = A_{0x} + \int s(x - w)\, F(w, \ldots)\, \mathrm{d}w, \tag{15}$$

where

$$\mathscr{D}_x A_{0x} = 0. \tag{16}$$

Applying this procedure to (13), the resulting equation relates the amplitude $A(x_i \ldots; y_j \ldots; z_k \ldots)$ in the Feynman manner, explicitly, to those amplitudes from which it could have arisen by a single operation of the interaction, which involves the nucleon $\psi(x)$. Thus the first term on the right-hand side, which in the integral form is

$$g\gamma(w) \int s(x_i - w)\, d(z_k - w)\, A(w \ldots; y_j \ldots; z_k^{-1} \ldots)\, \mathrm{d}w, \tag{17}$$

represents the possibility that the meson $\phi(z_k)$ has come from the point w, where it was emitted by a nucleon, which has itself propagated from w to the point x_i. A similar intepretation applies to the other terms and to the $\bar{\mathscr{D}}$ and $\mathscr{K}$ equations, which have not been written explicitly.

4. The integral equations

Given any amplitude $A(x_i \ldots; y_j \ldots; z_k \ldots)$ $(i = 1, \ldots, n; j = 1, \ldots, l; k = 1, \ldots, m)$, we have derived $n+l+k = N$ differential equations. Each of these selects out certain terms in which a particular particle is involved. The final task is to combine these N simultaneous differential equations into a single integral equation which is equivalent to the whole set of N equations. To accomplish this we state a third theorem.

Theorem 3. Let $D_1(x), D_2(x), \ldots$ be any set of commuting differential operators, whose inverses exist, and suppose

$$D_i(x_i)\, A(x_1 \ldots x_i \ldots x_N) = f_i(x_1 \ldots x_N) \quad (i = 1, \ldots, N), \tag{18}$$

where the functions f_i are given and assumed consistent with each other. From these equations can be derived other equations of the form

$$\left.\begin{aligned} &D_i D_j A = D_j D_i A = f_{ij} \\ \text{and} \quad &D_i D_j D_k A = f_{ijk}, \quad \text{etc.} \end{aligned}\right\} \tag{19}$$

Then the single integral equation which is equivalent to the whole set of N differential equations, (18), is

$$\begin{aligned} A = A_0 + \sum_i D_i^{-1} f_i - \frac{1}{2!} \sum_{i \neq j} D_i^{-1} D_j^{-1} f_{ij} \\ + \frac{1}{3!} \sum_{i \neq j \neq k} D_i^{-1} D_j^{-1} D_k^{-1} f_{ijk} - \ldots \\ + (-1)^{N-1} D_i^{-1} \ldots D_N^{-1} f_{1 \ldots N}, \end{aligned} \tag{20}$$

where

$$D_i A_0 = 0 \quad \text{for any } i. \tag{21}$$

The proof is given in the appendix.

Applying theorem 3 to the differential equations gives a single integral equation for each amplitude in terms of other amplitudes. We thus have an infinite set of coupled integral equations, analogous to the Fock equations between probability amplitudes in the 'old-fashioned' three-dimensional theory. For example

$$A(x; -; -) = A_0 + g\gamma(w) \int s(x-w)\, A(w; -; w)\, \mathrm{d}w \tag{22}$$

and

$$\begin{aligned} A(x; -; z) = A_0 + g\gamma(w) \int \mathrm{d}w [d(z-w)\, s(x-w)\, A(w; -; -) \\ + s(x-w)\, A(w; -; wz) + d(z-w)\, A(xw, w, -)] \\ - g^2 \gamma(w)\, \gamma(v) \iint \mathrm{d}w\, \mathrm{d}v [d(z-v)\, s(x-w)\, A(wv; v; w) \\ + d(z-v)\, s(x-w)\, s(w-v)\, A(v, -, w)]. \end{aligned} \tag{22'}$$

These equations can be interpreted graphically as was done for the differential equations. Note that the amplitude with N particles is related by these equations directly to an amplitude with $2N$ particles through a term of order g^N. This is a new and rather surprising feature of the theory.

No infinities occur in the kernels of these equations. An approximate integral equation for a single amplitude can be obtained by elimination of the other amplitudes. Such a procedure will produce divergences. Since all kernels are functions of the Feynman propagators and all integrations are over the whole of space-time, the divergent expressions have exactly the same form as in the S-matrix, and their finite parts can be defined in the same way. However, if the equations are solved by some method other than an expansion in g, all divergent terms of a certain type and order do not appear together, and it has not been shown that in this case the dropping of infinities can be interpreted logically as renormalizations.

It is interesting to note that although the development has been made in terms of the amplitudes A, we have at no stage used the fact that A is a matrix element. All our equations apply equally to the normal products defined by (11). The final equations (22) are thus an infinite set of integral equations for Heisenberg operators.

One of us (A. S.) would like to express his sincerest thanks to Professor R. E. Peierls, F.R.S., for the hospitality of the Mathematical Physics Department, Birmingham University, extended to him during the summer of 1953.

Appendix 1. Proof of theorem 2

Consider first

$$T(\psi(x), \overline{\psi}(y)) = \tfrac{1}{2}[\psi(x), \overline{\psi}(y)] + \tfrac{1}{2}\epsilon(x-y)\{\psi(x), \overline{\psi}(y)\}.$$

It follows that

$$\begin{aligned}\mathscr{D}_x T(\psi(x), \overline{\psi}(y)) &= g\gamma(x)\, T(\psi(x)\, \overline{\psi}(y)\, \phi(x)) + \mathscr{D}_x \epsilon(x-y)\{\psi(x), \overline{\psi}(y)\} \\ &= g\gamma(x)\, T(\psi(x)\, \overline{\psi}(y)\, \phi(x)) + \delta(x-y).\end{aligned}$$

The final δ-function comes from the relations

$$\frac{\partial \epsilon(x)}{\partial x_0} = 2\delta(x_0),$$

and the canonical commutation relation

$$\{\gamma_4 \psi(\mathbf{x}, t), \overline{\psi}(\mathbf{y}, t)\} = -\mathrm{i}\delta(\mathbf{x}-\mathbf{y}).$$

Similarly, the theorem may be proved for any other pair of operators.

Now consider $T(\psi(x), \mathscr{A}_i(x_i) \dots \mathscr{A}_n(x_n)) \equiv T$, where $\mathscr{A}_i$ may be either ψ, $\overline{\psi}$ or ϕ. We may assume without loss of generality that

$$x_{10} > x_{20} > \dots > x_{n0}.$$

Then, splitting time into periods which contain one operator $\mathscr{A}_i$ only, and summing,

$$\begin{aligned}T = \sum_{i=1}^{n} \mathscr{A}_1(x_1) \dots \mathscr{A}_{i-1}(x_{i-1})\, \theta[\tfrac{1}{2}(x_{i-1}+x_i) - x] \\ \times \theta[x - \tfrac{1}{2}(x_i + x_{i+1})]\, T(\psi(x), \mathscr{A}_i(x_i))\, \mathscr{A}_{i+1}(x_{i+1}) \dots \mathscr{A}_n(x_n),\end{aligned}$$

where

$$\theta(x) = \begin{cases} 1, & x_0 > 0, \\ 0, & x_0 < 0, \end{cases}$$

and θ-functions in the summation which involve x_0 or x_{n+1} are taken to be unity. Applying $\mathscr{D}_x$ to T gives the required result with the use of the theorem already established for just two operators and the relation

$$\frac{\mathrm{d}}{\mathrm{d}x_0}\theta(x) = \delta(x_0).$$

Appendix 2. Proof of theorem 3

It must be shown that operation with D_i on (20) reproduces the original differential equations. Suppose $A = A(x_1, x_2)$—a function of two variables only—and

$$D_1 A = f_1, \tag{18'a}$$

$$D_2 A = f_2. \tag{18'b}$$

Hence
$$D_1 D_2 A = h = D_2 f_1 = D_1 f_2. \tag{19'}$$

Then theorem 3 states

$$A = A_0 + D_1^{-1} f_1 + D_2^{-1} f_2 - D_1^{-1} D_2^{-1} h. \tag{20'}$$

Operating with D_1 on (20′) gives

$$D_1 A = f_1 + D_1 D_2^{-1} f_2 - D_2^{-1} h.$$

The last two terms cancel, by (19), so that

$$D_1 A = f_1,$$

which is (18a). Similarly for (18b). This argument can be generalized to a function of N variables using the identity given by the binomial expansion of $(1-1)^N = 0$. Hence the single integral equation (20) is equivalent to the set of N differential equations (18).

References

Bethe, H. A. & Salpeter, E. E. 1951 *Phys. Rev.* **84**, 1232.
Dancoff, S. M. 1950 *Phys. Rev.* **78**, 382.
Dyson, F. J. 1951 *Phys. Rev.* **82**, 428.
Dyson, F. J. 1953 *Phys. Rev.* **90**, 994.
Feynman, R. P. 1949 *Phys. Rev.* **76**, 749, 769.
Fock, V. 1932 *Z. Phys.* **75**, 622.
Gell-Mann, M. & Goldberger, M. L. 1952 *Phys. Rev.* **87**, 218.
Gell-Mann, M. & Low, F. E. 1951 *Phys. Rev.* **84**, 350.
Tamm, I. G. 1945 *J. Phys.* (*U.S.S.R.*) **9**, 449.
Wick, G. C. 1950 *Phys. Rev.* **80**, 268.

PHYSICAL REVIEW VOLUME 94, NUMBER 1 APRIL 1, 1954

Renormalization

P. T. MATTHEWS, *Department of Mathematical Physics, The University, Birmingham, England*

AND

ABDUS SALAM, *Government College, Lahore, Pakistan*

(Received October 7, 1953)

A theoretical justification for the infinite subtractions, which have to be made in the renormalization of the S matrix, is given along the lines suggested by Gupta and developed by Takeda. It is shown that this is equivalent to working with the renormalized field variables of Dyson, and that the method deals very simply with overlapping divergences and the "wave function" renormalization associated with external lines. It also gives directly Ward's identities and brings out their essential dependence on gauge invariance. The method is applied to free and bound electrons in electrodynamics and all renormalizable meson theories.

In the later sections the new method is related to the original method of Dyson; the Bethe-Salpeter equation is renormalized and closed forms are derived for the renormalization constants.

INTRODUCTION

THE general proof of the renormalization of the charge expansion of the S matrix of interacting fields[1] falls into three distinct parts. Firstly the number of types of infinity (primitive divergents) in the theory is determined. (If this number is finite the theory is renormalizable.) The second step is to define a subtraction procedure which removes these infinities. The third step is to provide a theoretical justification for these subtractions. A general outline of this proof, applied to electrodynamics and various meson theories, has already been given by us.[2] The purpose of the present review is to assume the results of the first two parts of the proof for any renormalizable theory, and to give, in detail, a treatment of the third part, which has previously presented the greatest difficulty. The central idea is to treat all divergences of the theory by means of infinite counter terms, as has always been done for the mass renormalization. This was first suggested by

[1] F. J. Dyson, Phys. Rev. 75, 1736 (1949).

[2] P. T. Matthews and Abdus Salam, Revs. Modern Phys. 25, 311 (1951).

Gupta[3] and the idea was developed along the lines presented here by Takeda.[4] The derivation of the required counter terms was given by Dyson.[5]

The great advantage of the present method is that it reduces to trivialities the two main problems of Dyson's original approach,[1] concerning the treatment of overlapping divergencies and of external lines.

This program occupies the first five sections. In Secs. VI and VII the present approach is related to the point of view adopted originally by Dyson[1] and followed in our general review.[2]

In the final sections the theory of the renormalization of the Bethe-Salpeter equation is given and closed forms are derived for the renormalization constants.

We would like to stress that this paper is intended mainly as a review and contributes little that is essentially new to the general theory. We feel that the subject matter is sufficiently important to justify its repetition in a complete and relatively simple form.

I. ELECTRODYNAMICS

The usual treatment starts with the Lagrangian:

$$\mathbf{L}=\mathbf{L}_0+\mathbf{L}_I, \tag{1.1}$$

$$\mathbf{L}_0(x)=-\tfrac{1}{4}\mathbf{F}_{\mu\nu}(x)\mathbf{F}_{\mu\nu}(x)-\tfrac{1}{2}[\nabla_\mu\mathbf{A}_\mu(x)]^2 -\psi^*(x)[\gamma_\mu\nabla_\mu+\kappa]\psi(x), \tag{1.2}$$

$$\mathbf{L}_I(x)=-ie\mathbf{A}_\mu(x)\psi^*(x)\gamma_\mu\psi(x)+\delta\kappa\psi^*(x)\psi(x). \tag{1.3}$$

Here κ is the observed mass and is related to the "bare" (or mechanical) mass κ_0 by

$$\kappa=\kappa_0+\delta\kappa. \tag{1.4}$$

However, e is the "bare" charge.[6] The Hamiltonian corresponding to this Lagrangian is

$$\mathbf{H}=\mathbf{H}_0+\mathbf{H}_I.$$

If one transform to the interaction representation (IR) by applying the transformation $U=\exp[i\mathbf{H}_0(t-t')]$ to the Schrödinger representation, defined at some fixed time t', the new equation of motion is

$$i\frac{\partial}{\partial t}\Psi)=\int H_I(\psi^*,\psi,A,e)d^3x\Psi), \tag{1.5}$$

where ψ^*, ψ, and A are interaction representation operators which annihilate and create bare particles with the observed mass. If one uses the subtraction procedure defined by Dyson, as extended by Salam to deal with overlapping divergences,[6a] one obtains finite matrix elements for the S matrix defined by this equation. The finite part of the theory is defined in such a way that it is both Lorentz and gauge invariant.

It is necessary finally to justify the apparently arbitrary dropping of the divergent terms. This is the part of the proof which will be given here in detail.

The subtraction procedure is defined in terms of infinite constants[7] $A(e)$ and $B(e)$, $C(e)$, and $L(e)$, the sum of the true divergences[6a] from electron, and photon, self-energy parts and vertex parts, respectively. We now take the essential step in the proof.[4] Define renormalized Heisenberg variables and observed charge by the equations

$$\psi=Z_2^{\frac{1}{2}}\psi_1,\quad \psi^*=Z_2^{\frac{1}{2}}\psi_1^*, \tag{1.6}$$

$$\mathbf{A}=Z_3^{\frac{1}{2}}\mathbf{A}_1,$$

$$e=Z_1Z_2^{-1}Z_3^{-\frac{1}{2}}e_1, \tag{1.7}$$

where

$$Z_1=1-L(e_1),\qquad Z_2=1+B(e_1),\quad Z_3=1+C(e_1). \tag{1.8}$$

ψ_1, ψ_1^*, and $\mathbf{A}_1$ are identical with the renormalized Heisenberg variables which have been discussed at length by Dyson,[5] but for our purposes they can be taken as defined by (1.6), (1.7), and (1.8). Expressing $\mathbf{L}$ in terms of renormalized Heisenberg variables, one obtains

$$\mathbf{L}\equiv\mathbf{L}_1\equiv\mathbf{L}_{10}+\mathbf{L}_{1I}, \tag{1.9}$$

where

$$\mathbf{L}_{10}=\mathbf{L}_0(\psi_1,\psi_1^*,\mathbf{A}_1) \tag{1.10}$$

is the same function of the new variables (provided we redefine the ∇A term which has no physical effect) and (see reference 24)

$$\mathbf{L}_{1I}=-ie_1[1-L(e_1)]\mathbf{A}_{1\mu}\psi_1^*\gamma_\mu\psi_1 -B(e_1)\psi_1^*[\gamma_\mu\nabla_\mu+\kappa]\psi_1+Z_2\delta\kappa\psi_1^*\psi_1 -\tfrac{1}{4}C(e_1)\mathbf{F}_{1\mu\nu}\mathbf{F}_{1\mu\nu}. \tag{1.11}$$

The corresponding Hamiltonian is

$$\mathbf{H}_1=\mathbf{H}_{10}+\mathbf{H}_{1I}.$$

Now transform to the renormalized interaction representation[8] (R.I.R.) by the transformation

$$U_1=\exp[i\mathbf{H}_{10}(t-t')]$$

---

[3] S. N. Gupta, Proc. Phys. Soc. (London) A64, 426 (1951).
[4] G. Takeda, Progr. Theoret. Phys. Japan 7, 359 (1952).
[5] F. J. Dyson, Phys. Rev. 83, 608 (1951).
[6] We denote by $\psi^*(\Psi^*)$ the quantity which appears elsewhere in the literature as $\bar{\Psi}$. $\psi^*=\psi\dagger\gamma_4$. [See J. Schwinger, Phys. Rev. 74, 1439 (1948)].
[6a] Reference 1, Secs. I–VI; Abdus Salam, Phys. Rev. 82, 217 (1951), Secs. II and III; and Phys. Rev. 84, 426 (1951).

[7] We use Dyson's notation, except that factors 2π (or $2\pi i$) have been absorbed in A, B, and C.
[8] Some care is required in deriving the Hamiltonian and the commutation relations in the renormalized interaction representation (R.I.R.). Firstly, all velocities $d\phi/dt$ must be eliminated from the Hamiltonian, which must be expressed in terms of the field variables and their canonical momenta, ϕ_1 and π_1, say, where

$$\pi_1=\frac{\partial L_1(\phi_1)}{\partial(d\phi_1/dt)}.$$

After transforming to the R.I.R. the interaction Hamiltonian is the same function of the new canonical variables $H_{1I}(\phi_1,\pi_1)$, where

$$\phi_1=U_1^{-1}\phi_{1s}U_1,$$

$$\pi_1=U_1^{-1}\pi_{1s}U_1,$$

where the suffix s denotes the Schrödinger representation. This shows that π_1 and ϕ_1 satisfy the equations of motion determined by H_{10}. Therefore π_1 is determined by the free Lagrangian only,

on the Schrödinger representation. The resulting equation of motion for the state vector is

$$i\frac{\partial}{\partial t}\Psi\rangle=\int H_{1I}(\psi_1^*,\psi_1,A_1,e_1)d^3x\Psi\rangle, \qquad (1.12)$$

where H_{1I} is essentially equal to $-L_{1I}$ apart from 'surface-dependent' terms which can safely be neglected since they cancel exactly with similar terms from the P brackets in the S matrix.[9]

The new field variables ψ_1^*, ψ_1, and A_1 are operators which annihilate and create a new type of bare particles which have not only the observed mass but also the observed charge. They will be referred to as renormalized bare particles. We now show that the new Hamiltonian contains just the right counter terms to cancel all divergences, provided $\delta\kappa$ is chosen so that

$$Z_2\delta\kappa=A(e_1), \qquad (1.13)$$

and that it leads directly to the finite results of the subtraction procedure.

It should be noted, firstly, that in this formulation no difficulty arises from self-energy insertions in external lines. If one works with ordinary bare particles in the interaction representation (I.R.) based on (1.3) these give rise to terms of the form [reference 1, Eq. (37)].

$$\psi'(p)=\psi(p)+S_F(p)\Sigma(p)\psi'(p)$$
$$=\psi(p)+S_F(p)B(e_1)S_F^{-1}(p)\psi(p), \qquad (1.14)$$

where, to obtain the second equality, the linear divergence of $\Sigma^*(p)$ has to be canceled by $\delta\kappa$ and the rest vanishes because it operates on an external particle. The remaining term is ambiguous because

$$S_F(p)\times[S_F^{-1}(p)\psi(p)]=0, \qquad (1.15)$$

but

$$[S_F(p)S_F^{-1}(p)]\times\psi(p)=\psi(p). \qquad (1.16)$$

The ambiguity is only removed if an adiabatic switching on and off of the charge is introduced to make (1.14) determinate.[10] However, working in the R.I.R. The counter terms are defined so that *all* the radiative corrections to an external line vanish identically. The analog of (1.14) is unambiguously

$$\psi_1'(p)=\psi_1(p). \qquad (1.17)$$

No adiabatic switching is required. The renormalized bare particles entering or leaving a scattering process can be identified with true free particles. All graphs containing insertions in external lines can be ignored.

Now consider the separation of internal divergences. When no overlaps occur, all graphs can be built up unambiguously by insertions of vertex and self-energy parts into all the vertices and lines of irreducible graphs. Whenever vertex parts or proper self-energy parts are inserted, one may also insert the appropriate new counter term of H_{1I} which, using (1.13) automatically cancels the infinities and carries out correctly Dyson's[1] subtraction procedure.

When overlaps occur, one must use Salam's prescription.[6] This is equivalent to Dyson's when no overlaps occur. Since this case has already been dealt with, Salam's prescription need only be applied explicitly to graphs involving overlaps, that is to say, in electrodynamics graphs.

Consider any self-energy graph—to be referred to as "the original" graph—of any degree of complexity, but built up of vertices at which only the main interaction term, $ieA_\mu\psi^*\gamma_\mu\psi$, is operating. If there exists a subpart of this graph, which is divergent, Salam's prescription is that one must subtract the infinite part (true divergence) of this subgraph, multiplied by the rest of the graph (reduced integral) obtained by shrinking to a single vertex the divergent subpart in the original graph (to be referred to as the "shrunk" vertex). If the graph has m nonoverlapping divergent subparts, with true divergences $D_1,\cdots,D_m$, then the true divergence of all the subparts together is defined to be $(-1)^mD_1D_2\cdots D_m$, and the reduced integral is the graph obtained by shrinking each of the m divergent subparts to a single vertex. A subtraction must be made for all possible ways of splitting the graph into divergent subparts and a reduced integral, including the case in which the subpart is the whole original graph. Finally, all self-energy graphs and their subtractions (up to some given order in the charge) must be summed.

Consider a particular way of splitting the original graph into a divergent subpart and a reduced integral. When the summation over all graphs is made, there occur a set of graphs all of which have the same reduced integral and the same type of subpart (vertex or self-energy), but in the set this part appears in all possible forms. Adding together the subtractions which must be made for this particular splitting of this set of graphs, one gets the reduced integral multiplied by the sum of all the true divergences (again up to the same given order in the charge) of all possible forms of the divergent subpart. But this is exactly the term which will arise from the Hamiltonian H_{1I} of (1.12) if the appropriate counter term is operating at the "shrunk" vertex of this set of graphs. This argument is readily extended to the case when the splitting of the original graph involves two or more nonoverlapping subparts, and leads to a subtraction term with two or more "shrunk" vertices at which the appropriate counter terms of H_{1I} are operating. Summing over all possible graphs given by H_{1I}, one sees that also when overlaps occur the sub-

namely

$$\pi_1=\frac{\partial L_{1\mathrm{R}}(\phi_1)}{\partial\dot{\phi}_1}.$$

Hence, ψ_1^*, ψ_1, and A_1 satisfy the ordinary free field commutation relations (with no Z factor), and are correctly normalized to annihilate and create particles. The last equation can be used to express H_{1I} in terms of velocities in the R.I.R., to give (1.12).

[9] P. T. Matthews, Phys. Rev. 76, 684 (1949).

[10] F. J. Dyson, Phys. Rev. 83, 608 (1951), Sec. X. See particularly the remark following Eq. (178).

traction procedure is automatically carried out by the Hamiltonian H.

The "subtraction procedure" is thus not a mutilation of the original theory by the arbitrary dropping of infinite terms, but a simple reinterpretation brought about by a change of normalization—(infinite) change of scale—of the charge and field variables. This is its justification, and any 'subtractions' which can be interpreted in this way will be referred to as renormalizations.

II. THE BOUND INTERACTION REPRESENTATION

For many problems it is convenient to split the electromagnetic field into the radiation field and an external part, due to external sources, which is assumed to be given classical function. For this purpose we replace A above by $A^{ex}+A$, where A now means the quantized radiation field. Following through exactly the same argument as before one obtains

$$\mathbf{L}_{10}=\mathbf{L}_0(\psi_1^*,\psi_1,\mathbf{A}_1^{ex}+\mathbf{A}_1) \tag{2.1}$$

and additional terms in the 'interaction.'

$$\mathbf{L}_{1I}^{ex}=-ie_1[1-B(e_1)]A_{1\mu}^{ex}\psi_1^*\gamma_\mu\psi_1 -\tfrac{1}{4}C(e_1)[2F_{1\mu\nu}^{ex}\mathbf{F}_{1\mu\nu}+F_{1\mu\nu}^{ex}F_{1\mu\nu}^{ex}]. \tag{2.2}$$

These will be referred to as the external interaction term plus three external counter terms. If the external field is weak, so that an expansion in terms of it is suitable, the renormalization argument goes through just as before. The three external counter terms being just those required to cancel the infinities, which involve the external interaction term. The last one makes finite the vacuum-to-vacuum expectation value of the S matrix which appears as an essential factor in the matrix element of any process if the external field can create pairs.[11]

Of more interest is the case of a strong external field, the outstanding example of which is of course the theory of the Lamb shift in which the electron under consideration is bound in the potential of the nucleus. For these problems we include the external interaction term in $H_{1\mathfrak{o}}$, and put only the external counter terms in H_{1I}. With this splitting of the Hamiltonian one can pass to the renormalized bound interaction representation (R.B.I.R.). This is similar to that introduced by Furry[12] in that the electron field operators annihilate and create electrons in bound states and the electron propagator S_F^{ex} is now the Feynman sum over bound states. It differs from Furry's bound interaction representation in that it contains counter terms with constants defined by the free I.R. Now any divergent graph (self-energy or vertex parts) in the R.B.I.R. will involve $S_F^{ex}(p)$ at least once. This can be replaced by the *exact* expression,

$$S_F^{ex}(p)=S_F(p)+S_F(p)\gamma A^{ex}(p)S_F(p) +S_F(p)\gamma A^{ex}S_F^{ex}(p)\gamma A^{ex}S_F(p). \tag{2.3}$$

In this way any divergent expression involving S_F^{ex} may be expressed as a divergent expression involving S_F plus a finite expression involving S_F^{ex} [the terms in S_F^{ex} are finite since at least two extra external lines have been introduced, corresponding to the two factors A^{ex} in the last term of (2.3)]. But since the first two terms in the right-hand side of (2.3) are the beginning of the expansion of S_F^{ex} in terms of the external field, the internal lines of all the divergent parts obtained in this way, are identical with those obtained in the renormalized free interaction representation. Since the infinities were canceled there by the counter terms, they will be canceled here too by the same counter terms to any order in the charge. Thus the R.B.I.R. gives directly finite results for any cross section or energy shift. This is the theoretical basis of the subtraction procedure adopted by Baranger[13] and by Kroll and Pollock.[14]

III. WARD'S IDENTITY

It has been shown above that a theory with divergences of the type which appear in electrodynamics can be renormalized even if no relation exists between the 'true divergences' of self-energy and vertex parts. However, the infinite constants have in fact been defined in such a way that the finite S-matrix elements given by the subtraction procedure are gauge invariant. This implies that if these finite elements can be derived from a new Lagrangian, this new Lagrangian must also be gauge invariant. Therefore the terms involving ψ_1 and either A_1 or ∇ in L_1 must combine to give an expression involving only

$$\nabla+ie_1A_1. \tag{3.1}$$

Since the original Lagrangian was gauge invariant, this is equivalent to the condition

$$e_1A_1=eA. \tag{3.2}$$

It follows that

$$Z_1=Z_2 \tag{3.3}$$

and hence that

$$L(e_1)=-B(e_1). \tag{3.4}$$

This is Ward's identity.[15] The above derivation brings out very clearly its direct dependence on the gauge invarance of the theory. This argument is due to Takeda.[4]

IV. MESON INTERACTIONS

It is clear that exactly the same argument can be applied to any renormalizable meson-nucleon interaction. Consider, for example, the pseudoscalar inter-

[11] R. P. Feynman, Phys. Rev. 76, 749 (1949). Abdus Salam, and P. T. Matthews, Phys. Rev. 90, 690 (1953).
[12] W. H. Furry, Phys. Rev. 81, 115 (1951).
[13] M. Baranger, Phys. Rev. 84, 866 (1951).
[14] N. M. Kroll and F. Pollock, Phys. Rev. 86, 876 (1952).
[15] J. C. Ward, Phys. Rev. 78, 182 (1950).

action of pseudoscalar mesons. All that is required is to replace the free photon terms in the Lagrangian by

$$\mathrm{L}_m=-\tfrac{1}{2}[(\nabla\phi)^2+\mu^2\phi^2], \tag{4.1}$$

and

$$\mathrm{L}_I=-ig\phi\psi^*\gamma_s\psi+\tfrac{1}{2}\delta\mu^2\phi^2+\delta\lambda\phi^4+\delta\kappa\psi^*\psi, \tag{4.2}$$

where the masses are already renormalized and the necessary contact interaction $\delta\lambda\phi^4$ has been introduced. Changing to renormalized variables as in the previous section, produces a Lagrangian which leads directly to the finite theory.

An even greater simplification is brought about by this approach in the interaction of spin-zero mesons with the electromagnetic field, where the overlapping of divergences can be very complicated. Here L consists of the free meson and photon terms, and

$$\mathrm{L}_I=ie'\mathrm{A}_\mu(\phi\nabla_\mu\phi^*-\nabla_\mu\phi\cdot\phi^*)-e^2\mathrm{A}^2\phi\phi^*+\delta\mu^2\phi\phi^*+\delta\lambda\phi^{*2}\phi^2. \tag{4.3}$$

e is supposed equal to e' (so that L is gauge invariant) but we distinguish, for the moment, between charges in the linear and bilinear terms. The S matrix based on this interaction is infinite, but can be made finite by subtractions which are defined in terms of infinite constants L, B, C, R given by Salam.[16] The next step is to define renormalized field variables, which introduce the required counter terms into the theory. It is not hard to see that we must take

$$\phi=Z_2^{\frac{1}{2}}\phi_1 \quad \mathrm{A}=Z_3^{\frac{1}{2}}\mathrm{A}_1 \tag{4.4}$$

where

$$\begin{aligned} Z_3&=1+B(e_1),\\ Z_2&=1+C(e_1). \end{aligned} \tag{4.5}$$

This gives the required self-energy counter terms. To cancel these in the main interaction terms and introduce the required factors in L and R, we must define

$$\begin{aligned} e'&=Z_3^{-\frac{1}{2}}Z_1Z_2^{-1}e_1,\\ e^2&=Z_3^{-1}Z_2^{-1}Z_4e_1^2, \end{aligned} \tag{4.6}$$

where

$$Z_1=1-L(e_1), \quad Z_4=1-R(e_1). \tag{4.7}$$

Then

$$\begin{aligned} L_{1I}=ie_1[1-L(e_1)]A_{1\mu}(\phi_1\nabla_\mu\phi_1^*-\nabla_\mu\phi_1\cdot\phi_1^*)\\ -e_1^2[1-R(e_1)]A_1^2\phi_1\phi_1^*+\delta\mu^2Z_2\phi_1\phi_1^*\\ +\delta\lambda Z_1^2\phi_1^{*2}\phi_1^2-\tfrac{1}{4}C(e_1)F_{1\mu\nu}F_{1\mu\nu}\\ -B(e_1)(\nabla_\mu\phi_1\nabla_\mu\phi_1^*+\mu^2\phi_1\phi_1^*). \end{aligned} \tag{4.8}$$

The subtractions defined by Salam,[6] whether overlapping divergences are involved or not, take place automatically when the S matrix is based on this Lagrangian, which thus provides the theoretical basis of the finite theory.

By this approach we have avoided all the complicated argument required when the Z factors are introduced and justified by the method of Salam[16] (following Dyson), or the ingenious, but far from perspicuous manipulations involved in the proof due to Ward.[17] Also, as above, no ambiguity arises from external lines.

Further, by the argument given in Sec. III, the renormalized Lagrangian is gauge invariant. Therefore

$$eA=e'A=e_1A_1. \tag{4.9}$$

Hence

$$Z_1=Z_2=Z_4. \tag{4.10}$$

It is clear that these considerations can be extended immediately to the combined interaction of three fields.

This completes our main purpose of providing, in a simple way, a theoretical justification for the infinite subtractions which occur in the renormalization of a divergent S matrix.

V. ALTERNATIVE FORMULATIONS

We have shown that the Lagrangian (1.9) leads to an S matrix in electrodynamics which is finite, term by term, in the charge expansion. The result of course does not depend on the particular split between the "free" and interaction parts. We may, alternatively, write exactly the same Lagrangian in the form

$$\mathrm{L}\equiv\mathrm{L}\dagger=\mathrm{L}_0\dagger+\mathrm{L}_I\dagger, \tag{5.1}$$

where

$$\mathrm{L}_0\dagger=-\tfrac{1}{4}Z_3\mathrm{F}_{1\mu\nu}\mathrm{F}_{1\mu\nu}-\tfrac{1}{2}Z_3(\nabla_\mu\mathrm{A}_{1\mu})^2-Z_2\psi^*(\gamma_\mu\nabla_\mu+\kappa)\psi, \tag{5.2}$$

$$\mathrm{L}_I\dagger=-ie_1Z_1\mathrm{A}_{1\mu}\psi_1^*\gamma_\mu\psi_1+Z_2\delta\kappa\psi_1\psi_1, \tag{5.3}$$

and define a new interaction representation[18] in terms of $L_0\dagger$ in which the variables are denoted by $\dagger$. In this representation the complete electron and photon propagators are infinite, that is S_{F1}', not S_F', is generated by making all possible self-energy insertions in a simple line. However, the simple propagator contains infinities coming from the Z factors in $L_0\dagger$. Thus $S_F\dagger(x)=Z_2^{-1}S_F(x)$, $D_F\dagger(x)=Z_3^{-1}D_F(x)$. Again the Z factors in $L_0\dagger$ gives rise to Z factors in the commutation relations of $\psi_1^*\dagger$, $\psi_1\dagger$, and $A_1\dagger$ so they are not correctly normalized for annihilators and creators of bare particles. To avoid this and to get graphs with simple propagators S_F rather than $S_F\dagger$ we express the interaction in terms of $Z_3^{\frac{1}{2}}A_1\dagger$, $Z_2^{\frac{1}{2}}\psi_1^*\dagger$ and $Z_2^{\frac{1}{2}}\psi_1\dagger$. Or equivalently we continue to work with unrenormalized field variables and introduce only the charge renormalization (1.7). This gives the original Lagrangian (1.1) but with the interaction expressed in terms of the renormalized charge

$$L_I=-ie_1Z_2^{-\frac{1}{2}}Z_1Z_2^{-1}A_\mu\psi^*\gamma_\mu\psi+\delta\kappa\psi^*\psi, \tag{5.4}$$

which again must lead to a finite S matrix when expressed as a power series in e_1. This is the formulation

[16] Abdus Salam, Phys. Rev. 86, 731 (1952).

[17] J. C. Ward, Phys. Rev. 84, 897 (1951).

[18] S. Kamefuchi and H. Umezawa, Progr. Theoret. Phys. Japan 7, 399 (1952).

given by Dyson,[1] but established quite differently by him.

VI. DYSON'S RELATIONS

For completeness we now derive from our definitions, the relations by which the Z factors were defined by Dyson.[1]

By definition

$$-\tfrac{1}{2}S_F(y-y^1)=\langle T(\psi(y),\psi^*(y^1))\rangle_0, \tag{6.1}$$

where T is the chronological product as defined by Wick.[19] It follows immediately from the graphical definition of S_F', as the electron propagator modified by all possible self-energy insertions, that

$$-\tfrac{1}{2}S_F'(y-y^1)=\langle T(S,\psi(y),\psi^*(y^1))\rangle_0/\langle S\rangle_0, \tag{6.2}$$

where $\langle\cdots\rangle_0$ is the bare vacuum expectation value and

$$T(S,\psi(y),\psi^*(y^1))=\sum_n\frac{(-i)^n}{n!}\int\cdots\int dz_1\cdots dz_n T(H(z_1)\cdots H(z_n),\psi(y),\psi^*(y^1)). \tag{6.3}$$

By the argument given by Low and Gell-Mann[20] and using their definition of the true vacuum $\Psi_0\rangle$, it follows that

$$-\tfrac{1}{2}S_F'(y-y^1)=\langle\Psi_0 T(\psi(y),\psi^*(y^1))\Psi_0\rangle. \tag{6.4}$$

By an exactly similar argument, only replacing all operators by the corresponding renormalized operators, the finite propagator is

$$-\tfrac{1}{2}S_{F1}'(y-y^1)=\langle\Psi_0 T(\psi_1(y),\psi_1^*(y^1))\Psi_0\rangle. \tag{6.5}$$

Using the definitions of ψ_1 and ψ_1^*, it follows that

$$S_F'(e)=Z_2S_{F1}'(e_1). \tag{6.6}$$

Similarly,

$$D_F'(e)=Z_1D_{F1}'(e_1). \tag{6.7}$$

Let $\Lambda_1(e_1)$ be the sum over all vertex parts obtained from H_{1I} in the R.I.R. Consider any vertex part which does not contain any self-energy part. The effect of all self-energy insertions, including their counter terms, is given by writing $S_{F1}'(e_1)$ and $D_{F1}'(e_1)$ for the lines. The vertex counter term is included by multiplying by $1-L(e_1)(=Z_1)$ at each vertex. If the part has $2n+1$ vertices, there are $2n$ electron lines and n photon lines. Using relations (1.7), (6.6), and (6.7), and summing over all such vertex parts, gives

$$Z_1^{-1}\Lambda_1(e_1)=\Lambda(e), \tag{6.8}$$

where $\Lambda(e)$ is the sum over all vertex parts given by H_I in the I.R. Now

$$\Gamma(e)=\gamma+\Lambda(e)=Z_1^{-1}[Z_1\gamma+\Lambda_1(e_1)]=Z_1^{-1}\Gamma_1(e_1). \tag{6.9}$$

(6.7), (6.8), and (6.9) are the relations which Dyson[1] uses to *define* the Z factors. It should be noted that they are not required for the general proof of renormalization as presented here.

VII. THE BETHE-SALPETER EQUATION

The derivation of the Bethe-Salpeter[21] equation for pseudoscalar coupling given by Low and Gell-Mann[20] can be repeated step by step using renormalized quantities throughout, to derive the renormalized equation, which contains no infinities.

Alternatively, one can start from the unrenormalized form, which can be written symbolically as[22]

$$K=S_F'S_F'-\int S_F'S_F'GK. \tag{7.1}$$

Define the renormalized two-particle propagator

$$K_1(x_1,x_2,x_3,x_4)=\langle\Psi_0 T(\psi_1(x_1),\psi_1(x_2),\psi_1^*(x_3),\psi_1^*(x_4))\Psi_0\rangle. \tag{7.2}$$

Then, immediately from the definitions of ψ_1, etc.,

$$K=Z_2^2K_1. \tag{7.3}$$

The interaction function G can be derived from the sum of all irreducible graphs with just four external nucleon lines, by replacing the lines and vertices by S_F', Δ_F' and Γ_5. A graph with $2n$ vertices has n meson lines and $2n-2$ nucleon lines. Therefore, by (6.7), (6.8), and (6.9),

$$G(e)=Z_2^{-2}G_1(e_1). \tag{7.4}$$

Substituting into (7.1), all the Z factors cancel and

$$K_1=S_{F1}'S_{F1}'-\int S_F'S_F'G_1K_1,$$

which is the finite equation for the renormalized propagator. Similarly, the renormalized "wave function" is

$$\chi_{n1}(x_1,x_2)=\langle\Psi_0 T(\psi_1(x_1),\psi_1^*(x_2))\Psi_n\rangle=Z_2^{-1}\chi_n(x_1,x_2),$$

and again, on substitution into the B-S equation,[22] all Z factors cancel, leaving the renormalized equation:

$$\chi_1=-\int S_F'S_F'G_1\chi_1.$$

It is important for this argument that no divergences are produced by the iteration of the equation for K.

VIII. CLOSED FORM OF THE THEORY

In this final section the infinite constants introduced in (1.6)–(1.8) are expressed in terms of renormalized field variables.

[19] G. C. Wick, Phys. Rev. 80, 268 (1950).
[20] M. Gell-Mann and F. E. Low, Phys. Rev. 84, 350 (1951).
[21] H. A. Bethe and E. E. Salpeter, Phys. Rev. 84, 1232 (1951).
[22] Reference 20, Eq. (23).
[22] Reference 20, Eq. (37).

Returning to S_F', we have the general relation:

$$S_F'(y-y^1)=S_F(y-y^1)+\int\int dzds^1 S_F(y-z)\Sigma(z-z^1)S_F(z^1-y^1). \quad (8.1)$$

Compare this with (6.1) and (6.2). The first term on the right-hand side of (8.1) is given by putting $n=0$. The second term is got by combining $\psi(y)$ with any $\psi^*(z)$ and $\psi^*(y^1)$ with any $\psi(z^1)$ ($z^1\neq z$) in the T bracket, which can be done in $n(n-1)$ ways in the nth term. There are additional nonvanishing terms in which $\psi(y)$ and $\psi^*(y^1)$ are combined with the $\psi^*(z)$ and $\psi(z)$ of a single $\delta\kappa$ term, which may be done in n ways in the nth term. One thus obtains an expression for $\Sigma(z-z^1)$, namely

$$\Sigma(z-z^1)=\langle T(S,f(z),f^*(z^1))\rangle_0/\langle S\rangle_0-\tfrac{1}{2}i\delta\kappa\delta(z-z^1) \quad (8.2)$$

$$=\tfrac{1}{2}[\langle\Psi_0 T(\mathfrak{f}(z),\mathfrak{f}^*(z^1))\Psi_0\rangle-i\delta\kappa\delta(z-z^1)], \quad (8.3)$$

where

$$\mathfrak{f}=ig\phi\gamma_5\psi-\delta\kappa\psi,\qquad \mathfrak{f}^*=ig\phi\psi^*\gamma_5-\delta\kappa\psi^*, \quad (8.4)$$

which, for pseudoscalar interaction, is the analog of the current in the equations of motion for ψ and ψ^*, respectively. To obtain (7.3) we have again used the argument of Low and Gell-Mann.[20]

Alternatively, working with renormalized variables

$$\Sigma_1(z-z^1)=\tfrac{1}{2}[F_1(z-z^1)+iZ_2\delta\kappa-B(e_1)(\gamma_\mu\nabla_\mu+\kappa))\delta(z-z^1)], \quad (8.5)$$

where

$$F_1(z-z^1)=\langle\Psi_0 T(\mathfrak{f}_1(z),\mathfrak{f}_1^*(z^1))\Psi_0\rangle, \quad (8.6)$$

and

$$\mathfrak{f}_1=igZ_1\phi_1\gamma_5\psi_1+B(e_1)(\gamma_\mu\nabla_\mu+\kappa)\psi_1-Z_2\delta\kappa\psi_1. \quad (8.7)$$

Expressing (8.5) in momentum space,

$$\Sigma_1(p)=\tfrac{1}{2}[F_1(p)+iZ_2\delta\kappa+B(e_1)(p-i\kappa)] \quad (8.8)$$

where

$$p=p_\mu\gamma_\mu. \quad (8.9)$$

But the infinite constants B and $\delta\kappa$ have been defined so that

$$\Sigma_1(i\kappa)=0, \quad (8.10)$$

and

$$[\partial\Sigma_1/\partial p]_{i\kappa}=0. \quad (8.11)$$

Therefore

$$Z_2\delta\kappa=-iF_1(i\kappa), \quad (8.12)$$

and

$$B(e_1)=-[\partial F_1/\partial p]_{i\kappa}. \quad (8.13)$$

Similar expressions can be obtained for $Z_3\delta\mu^2$ and $C(e_1)$ by considering Δ_{F1}'.

Again, from a consideration of the graphs, it can be seen that the finite part of all vertex parts is

$$\Lambda_1(x,y,z)=\langle T(S_1,f_1^1(x)f_1^{*1}(y),\times j_1^1(z))\rangle_0/\langle S_1\rangle_0-\gamma L \quad (8.14)$$

where

$$f_1^1=ig_1Z_1\phi_1\gamma_5\psi_1, \quad (8.15)$$

$$j_1^1=Z_1\psi_1^*\gamma_5\psi_1. \quad (8.16)$$

Thus, adding the term from the single vertex,

$$\Gamma_1(x,y,z)=G_1(x,y,z)+\gamma_5(1-L), \quad (8.17)$$

where

$$G_1(x,y,z)=\langle\Psi_0 T(\mathfrak{f}_1^1(x),\mathfrak{f}_1^{*1}(y),\mathfrak{j}_1^1(z))\Psi_0\rangle.$$

In momentum space,

$$\Gamma_1(p,p^1)=G(p,p^1)+\gamma_5(1-L). \quad (8.18)$$

Also, since

$$\Gamma_1(i\kappa,i\kappa)=\gamma_5, \quad (8.19)$$

$$\gamma_5 L=G_1(i\kappa,i\kappa). \quad (8.20)$$

Relations similar to (8.12) and (8.13) were first obtained by Kallen.[24] With the use of these expressions it is possible to state the renormalized theory, without mention of power series expansions. Thus the Lagrangian is (1.9)–(1.11). The infinite constants appearing in the Lagrangian are defined by the implicit relations (8.12), (8.13), (8.20), etc.

We do not wish to suggest that this is necessarily a finite theory independent of charge expansion. Dyson's analysis[25] of the divergences into primitive divergents depends essentially on the charge expansion. If this expansion is not absolutely convergent, there is no reason to suppose that the S-matrix elements, obtained by some method other than a charge expansion, would have divergences restricted in the same way and removable by the same renormalizations.[26]

One of us (A. S.) would like to thank Prof. R. E. Peierls for the hospitality of Birmingham University extended to him during the summer of 1953. We also acknowledge helpful conversations with Dr. J. G. Valatin.

[24] G. Kallen, Helv. Phys. Acta **25**, 417 (1952). If Σ_1 is defined by the relation,

$$\langle\Psi_0 T(\psi,\psi^*)\Psi_0\rangle=-\tfrac{1}{2}(S_F+S_F\Sigma_1 S_F),$$

and the infinite constants B and $\delta\kappa$ are defined by (8.10) and (8.11), then Eqs. (8.12) and (8.13) can be derived without the use of power series by methods employed by P. T. Matthews and Abdus Salam, Proc. Roy. Soc. (London) **A221**, 128, Sec. III.

[25] Reference 1, Sec. V.

[26] See, for example, G. Feldman, Proc. Roy. Soc. (London) (to be published).

Reprinted from the *Proceedings of the Cambridge Philosophical Society*
Volume 51, Part 1, pp. 103–112, 1955
© Cambridge University Press

THE BOUND STATE REPRESENTATION

By J. HAMILTON and A. SALAM

Received 17 June 1954

ABSTRACT. Eigenvalues and eigenvectors of the total Hamiltonian in field theory are determined by the method of direct diagonalization, using canonical transformations. The difficulties of this method are discussed in detail.

1. Two main approaches to the problem of bound states in quantum field theory have been made: (*a*) methods based on the approximation of Tamm and Dancoff (1, 6, 11), (*b*) direct derivation of a bound state equation by Salpeter and Bethe (9). The original Tamm–Dancoff method suffered from serious difficulties arising from the vacuum self energy, and Dyson (3) proposed a modified form to avoid these defects. However, Taylor (12) has shown that because the modified equations are not relativistically covariant, the particular subset of processes which the method selects cannot be made to yield finite results by charge and mass renormalization. The equation of Salpeter and Bethe can be derived either directly, as Gell-Mann and Low (4) did, or by means of Feynman amplitudes following Matthews and Salam (8). Gell-Mann and Low's method gives a kernel which is expressed as a power series whose convergence has not been proved. In Matthews and Salam's method the equation contains a remainder whose size has not been investigated.

In this paper we wish to reconsider the direct method of obtaining eigenvalues and eigenstates of the Hamiltonian by using a canonical transformation. Consider a fermion field and a boson field whose behaviour is described by a Hamiltonian $H = H_0 + H_1$, H_1 being the interaction part. A suitable example is the interacting nucleon and uncharged pseudoscalar meson fields (with a pseudoscalar interaction). $\psi(\mathbf{x}, t)$, $\overline{\psi}(\mathbf{x}, t)$ (where $\overline{\psi} = \psi^\dagger \gamma^4$) are field operators describing the nucleons and $\phi(\mathbf{x}, t)$ is the meson field operator. On any time plane $t = t_0$ the commutation relations are

$$\left.\begin{aligned} [\phi(\mathbf{x}, t_0), \psi(\mathbf{x}', t_0)]_- = [\phi(\mathbf{x}, t_0), \overline{\psi}(\mathbf{x}', t_0)]_- &= 0, \\ [\overline{\psi}(\mathbf{x}, t_0), \overline{\psi}(\mathbf{x}', t_0)]_+ = [\psi(\mathbf{x}, t_0), \psi(\mathbf{x}', t_0)]_+ &= 0, \\ [\overline{\psi}_\beta(\mathbf{x}, t_0), \psi_\alpha(\mathbf{x}', t_0)]_+ &= \gamma^4_{\alpha\beta} \delta^3(\mathbf{x} - \mathbf{x}'), \\ [\phi(\mathbf{x}, t_0), \phi(\mathbf{x}', t_0)]_- &= \frac{\hbar}{i} \delta^3(\mathbf{x} - \mathbf{x}'). \end{aligned}\right\} \tag{1}$$

Fourier analysis gives

$$\left.\begin{aligned} \psi_\sigma(\mathbf{x}, t_0) &= \sum_{E_m > 0} a_m(t_0)\, v_{m\sigma}\, e^{i\mathbf{m}.\mathbf{x}} + \sum_{E_m < 0} \tilde{b}_m(t_0)\, v_{m\sigma}\, e^{i\mathbf{m}.\mathbf{x}}, \\ \overline{\psi}_\sigma(\mathbf{x}, t_0) &= \sum_{E_m > 0} \tilde{a}_m(t_0)\, v^*_{m\sigma}\, e^{i\mathbf{m}.\mathbf{x}} + \sum_{E_m < 0} b_m(t_0)\, v^*_{m\sigma}\, e^{i\mathbf{m}.\mathbf{x}}, \\ \phi(\mathbf{x}, t_0) &= \sum_{\mathbf{k}} \sqrt{\left(\frac{\hbar^2}{2\epsilon_k}\right)} [q_k(t_0) + \tilde{q}_k(t_0)]\, e^{i\mathbf{k}.\mathbf{x}}, \end{aligned}\right\} \tag{2}$$

where $E_m = \pm \hbar c \sqrt{(K^2 + m^2)}$, $\epsilon_k = \hbar c \sqrt{(\mu^2 + k^2)}$ and $v_{m\sigma}$ are spinors. It is characteristic of the canonical formulism that $\tilde{a}, a; \tilde{b}, b; \tilde{q}, q$ can be considered as creation and

annihilation operators at time t_0 for free nucleons, anti-nucleons and mesons, respectively†. $\tilde{a}, a$ will also be used to describe collectively any creation or annihilation operator respectively. These operators satisfy the commutation relations

$$\left.\begin{aligned} &[a_m(t_0), \tilde{a}_{m'}(t_0)]_+ = \delta_{mm'}, \quad [b_m(t_0), \tilde{b}_m(t_0)]_+ = \delta_{mm'}, \\ &[a_m(t_0), a_{m'}(t_0)]_+ = [a_m(t_0), \tilde{b}_m(t_0)]_+ = [a_m(t_0), b_m(t_0)]_+ = 0, \\ &[q_k(t_0), \tilde{q}_{k'}(t_0)]_- = \delta_{kk'}, \quad [q_k(t_0), a_m(t_0)]_- = 0, \\ &[q_k(t_0), b_m(t_0)]_- = [q_k(t_0), \tilde{a}_m(t_0)]_- = 0, \quad \text{etc.} \end{aligned}\right\} \tag{3}$$

The energy eigenvalues $\mathscr{E}_i$ and the Heisenberg eigenstates Φ_i are obtained from the equation

$$[H_0(a) + H_1(a)]\,\Phi_i = \mathscr{E}_i \Phi_i. \tag{4}$$

Here $H_1(a) = ig \int \bar{\psi}(a)\, \gamma^5 \psi(a)\, \phi(a)\, d^3x$, and a is used to show that H_0 and H_1 are written in terms of the operators $a(t_0)$, etc. defined by (2).

Define operators N^+, N^-, N^0 by the relations

$$N^+ = \Sigma \tilde{a}_m a_m, \quad N^- = \Sigma \tilde{b}_m b_m, \quad N^0 = \Sigma \tilde{q}_m q_m.$$

N^+, N^-, N^0 have as eigenvalues the total number of nucleons, anti-nucleons and bosons. A product of creation and annihilation operators will be called a D-product if it has an equal number of creation and annihilation operators for each kind of particle (i.e. for nucleons, anti-nucleons and mesons). The operators in a D-product are to be ordered so that all annihilation operators stand to the right of creation operators. For example $\tilde{a}_l \tilde{b}_m a_p b_r$ is a D-product. Clearly all D-products commute with N^+, N^-, N^0.

Consider a unitary operator T which makes H diagonal in the number of particles of each kind. The new variables A are related to the Heisenberg variables a by

$$A(t_0) = T a(t_0) \tilde{T}. \tag{5}$$

T has the properties $\tilde{T}T = T\tilde{T} = I$ and

$$\begin{aligned} H_0(a(t_0)) + H_1(a(t_0)) &= \tilde{T}[H_0(A(t_0)) + H_1(A(t_0))]\,T \\ &= H_0(A(t_0)) + V(A(t_0)). \end{aligned} \tag{6}$$

Here $V(A)$ is a sum of all D-products in the new variables A (except $\tilde{A}A$, $\tilde{B}B$, $\tilde{Q}Q$), and has the form

$$V(A) = \Sigma C_{lmnr} \tilde{A}_l \tilde{A}_m A_n A_r + \Sigma D_{lmnr} \tilde{A}_l \tilde{Q}_m A_n Q_r + \ldots. \tag{7}$$

The coefficients C_{lmnr} secure the conservation of the momenta of the particles concerned but not of the energy. By definition

$$[V(A), N^+(A)]_- = [V(A), N^-(A)]_- = [V(A), N^0(A)]_- = 0. \tag{8}$$

The new representation defined by (5) will be called the bound state representation. The bound state representation (B.S.R.) operators $A(t_0)$ obey the same commutation relations (3) as the corresponding Heisenberg operators $a(t_0)$; but the time dependence of the B.S.R. operators will differ from the time dependence of the Heisenberg operators. Using (6), we have

$$\frac{\partial A(t_0)}{\partial t_0} = \frac{i}{\hbar}[H_0(a) + H_1(a), A(t_0)]_- = \frac{i}{\hbar}[H_0(A) + V(A), A(t_0)]_-. \tag{9}$$

† Splitting into creation and annihilation operators can be performed in a covariant manner on a general space-like surface, and the splitting does not depend on the surface (cf. Schwinger (11)). We shall for convenience work on the time plane t_0.

It is clear that T must depend on time if (5) is to remain true for any time, and it is easy to see that the correct time dependence of $T(t)$ is given by

$$i\hbar\frac{\partial T(t)}{\partial t} = [H_1(A) - V(A)]\,T(t). \tag{10}$$

The form chosen for V ensures that it acts as the potential operator. The real reason for using it, however, lies in the fact that there is a close correspondence between the eigenvectors of H_0 and those of $H_0 + V$.

The relation between the eigenstates of H_0 and $H_0 + V$. Let

$$|\,t_0\rangle_0, \quad \Psi'_l = \bar{a}_l(t_0)\,|\,t_0\rangle_0, \quad \Psi'_{lm} = \frac{1}{\sqrt{2}}\bar{a}_l\bar{a}_m\,|\,t_0\rangle_0, \quad \text{etc.},$$

be the (normalized) eigenstates of the free Hamiltonian $H_0(a(t_0))$. Defining new state functions

$$\Psi_0(t_0) = T\,|\,t_0\rangle_0, \quad \Psi_l(t_0) = T\Psi'_l(t_0), \quad \Psi_{lm}(t_0) = T\Psi'_{lm}(t_0)$$

we call Ψ_0, Ψ_l, Ψ_{lm}, etc., the 'vacuum', 'one-particle' and 'two-particle' states in the new representation.

(*a*) *The vacuum state and the single-particle states.* Since $a_p\,|\,t_0\rangle_0 = 0$ by definition, we infer that

$$A_p\Psi_0 = 0. \tag{11}$$

Thus

$$(H_0(A) + V(A))\,\Psi_0 = 0, \tag{12}$$

and therefore Ψ_0 is a simultaneous eigenstate of $H_0(A)$ and $H_0(A) + V(A)$. The relations (11) and (12) hold always if they are true for one time, because, for example,

$$[H_0(A) + V(A), A_p]_-\,\Psi_0 = -A_p(H_0(A) + V(A))\,\Psi_0 = 0,$$

so

$$O = \frac{\partial A_p}{\partial t}\Psi_0 = \frac{\partial^2 A_p}{\partial t^2}\Psi_0, \quad \text{etc.}$$

Thus (11) can be taken as the definition of the vacuum state. This does not guarantee that the vacuum is also the eigenstate of least energy of the total Hamiltonian, but we will consider the latter assumption to be justified physically.

Ψ_l like Ψ_0 is a simultaneous eigenstate of H_0 and $H_0 + V$. It is not suggested that the eigenvalues of H_0 and $H_0 + V$ for these two states are identical. V may contain an unobservable vacuum self energy term which is ignored in (7).

(*b*) *Two-particle states.* It is important to note that the two-particle states are not eigenstates of $H_0 + V$. However, they have the simple property

$$V\Psi_{lm} = \sum_{j,k} F_{lmjk}\Psi_{jk},$$

where the F_{lmjk} are coefficients; thus the V operator converts two-particle states into a sum of two-particle states, without any admixture of three- or many-particle states. We believe this is the most that can be hoped for as an immediate diagonalization of the total Hamiltonian; but it should indeed be possible to make H diagonal to this extent, so on physical grounds we expect a unitary operator T with the required properties (equations (6) and (5)) to exist. Given V it is comparatively easy to set up an integral equation for the deuteron, for example. The deuteron state function would be written as a sum of two-particle states Ψ_{lm} alone. The equation of Salpeter

and Bethe (in the form given by Gell-Mann and Low) also gives a relation between two-particle states, but their equation contains two times while ours has only one time.

2. In this section the existence of the transformation operator T will be investigated. We shall show that this operator exists in the form of a series expansion. Let

$$T = \sum_{n=0} T_n, \tag{13}$$

where T_n contains g^n (g being the coupling constant). If we write $T_n = X_n + iY_n$ where X_n, Y_n are Hermitian, the unitary conditions $T\bar{T} = 1$ and $\bar{T}T = 1$ give, on picking out the coefficients of g^n,

$$\left.\begin{aligned} X_1 &= 0, \\ 2X_2 &= -(X_1^2 + Y_1^2), \\ 2X_3 &= i(Y_1X_2 - X_2Y_1) + i(Y_2X_1 - X_1Y_2) - (X_1X_2 + X_2X_1) - (Y_1Y_2 + Y_2Y_1). \end{aligned}\right\} \tag{14}$$

The general equation in (14) relates X_n to X_{n-1}, X_{n-2}, ..., Y_{n-1}, Y_{n-2}, ... The Y_n are not determined by the unitary condition.

Suppose

$$V = V_2 + V_4 + \ldots,$$

where V_{2n} contains g^{2n}, then (6) gives

$$\bar{T}(H_0 + H_1)\,T = H_0 + V_2 + V_4 + \ldots, \tag{15}$$

so on selecting the coefficient of g^n

$$\begin{aligned} i[H_0, Y_n]_- + (H_0X_n + X_nH_0) + i[H_1, Y_{n-1}]_- + (H_1X_{n-1} + X_{n-1}H_1) \\ + (\text{terms in } X_{n-2}, Y_{n-2}, \ldots, \text{etc.}) \end{aligned} = \begin{cases} 0, & n \text{ odd} \\ V_n, & n \text{ even} \end{cases}. \tag{16}$$

Thus (14) and (15) relate X_n, Y_n to X_{n-1}, X_{n-2}, ...; Y_{n-1}, Y_{n-2}, ...; V_s, V_{s-2}, ..., where $2s \leqslant n$. The equation for Y_n can be written

$$i[H_0, Y_n]_- = R_n, \tag{17}$$

where R_n is Hermitian. (17) is an equation determining Y_n as a function of the bound state representation variables on the time plane $t = t_0$ (cf. (2) and (3)), and it can be solved by using as auxiliary variables the interaction representation operators $\hat{A}(t)$ which coincide with the bound state representation operators $A(t_0)$ when $t = t_0$.

$$Y_n = \lim_{\gamma \to 0} \frac{1}{\hbar} \int_{-\infty}^{t_0} \hat{R}_n(t)\, e^{-\gamma(t_0 - t)}\, dt \tag{18}$$

is a solution of (17)† and the integral is a function of the variables $A(t_0)$, the interaction representation being used merely because it gives a convenient expression for the solution. When R_n has matrix elements connecting states of the same free particle energy (18) does not give a unique answer, and in this case it is necessary to refrain from taking the limit ($\gamma \to 0$) until manipulations with Y_n (such as multiplication by other operators) has been completed. The use of γ is then seen to be similar to the device of the small imaginary part due to Feynman—it enables us to avoid the singularity in a well-defined way. For elements of R_n which do not conserve free particle energy the factor $e^{-\gamma(t_0 - t)}$ merely has the effect of avoiding any dependence on the lower limit

† The solution of (17) is arbitrary to the extent that multiples of H_0 or of any variable commuting with H_0 can be added to the expression (18). We ignore this freedom and use (18) always.

of the integration. It must be emphasized that the use of the interaction representation in (18) does not imply that the time dependence of $A(t)$ (or any function of the $A(t)$ such as $Y_n(t)$) is of the free particle form; equation (9) gives the time dependence of these variables.

The choice of V_n. Write (16) as

$$i[H_0, Y_n]_- = \begin{cases} U_n & n \text{ odd}, \\ U_n + V_n & n \text{ even}. \end{cases} \tag{19}$$

U_n is determined by $U_{n-1}, U_{n-2}, \ldots, V_{n-2}, V_{n-4}, \ldots$, but so far the only condition on V_n is (15). We suggest that the 'natural' choice of V_n would be obtained by splitting U_n into a part which is diagonal in the numbers of particles $(U)_D$ (i.e. a sum of D-products), and a non-diagonal part $(U_n)_{\mathrm{ND}}$. (19) can then be written

$$i[H_0, Y_n]_- = (U_n)_{\mathrm{ND}}, \tag{20}$$

so that

$$V_n = -(U_n)_D. \tag{21}$$

(20) and (21) then determine V and T uniquely, and we call this the *natural choice*. For electrodynamics it gives the Coulomb field for V_2, and would thus give accurate results. The natural choice makes Y non-diagonal (in particle numbers) so the diagonal matrix elements of T and $\tilde{T}$ are identical.

There are no mathematical reasons why with V given by (21) the eigenvalues of $H_0 + V$ should be identical with those of H. Arguments based on the unitary character of T may well be misleading, as we shall see in § 4.

There are, however, some obvious criteria which the V and T matrices should satisfy: (1) V must have the correct covariance properties; (2) V and T should be divergence free; and, most important of all, (3) V and T should have power series expansions which are guaranteed to converge.

In spite of the basic character of criterion (3), in practice it is condition (2) about the divergence-free character of V and T which is the hardest to satisfy in field theory. Just to avoid the discussion of divergences, we illustrate our procedure for a finite Hamiltonian, in the next section. We shall set up an equation for the deuteron by using a finite form of the total Hamiltonian, proposed by Dyson (2). Dyson's method transforms the Hamiltonian by a preliminary unitary transformation which splits the low and high frequency parts. The transformed Hamiltonian contains convergence factors such that any further manipulations (e.g. those required to obtain V) do not produce any new divergence. In § 3 this Hamiltonian is used together with our natural choice of V.

3. *The deuteron equation.* A possible starting-point for deriving an equation of the deuteron is the renormalized Schrödinger equation which Dyson (2) obtains from the *intermediate representation*. The intermediate representation state function $\Phi_g(t)$ and operators $A_g(t)$ are related to the interaction representation state function $\Psi(t)$ and operators $A(t)$ by the unitary transformation.

$$\begin{aligned} \Psi(t) &= S_g(t)\,\Phi_g(t), \\ A_g(t) &= S_g(t)\,A(t)\,S_g(t), \end{aligned} \tag{22}$$

where

$$S_g(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{-i}{\hbar}\right)^n \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \ldots \int_{-\infty}^{t} dt_n\, P(H(t,t_1), H(t,t_2) \ldots H(t,t_n)). \tag{23}$$

Apart from counter terms which have to be added for renormalization,

$$H(t,t') = g(t-t')\,H_1(t'),$$

where H_1 is the interaction Hamiltonian of § 1 expressed in terms of interaction representation variables and the renormalized coupling constant. $g(a)$ is a real function of a positive variable a such that

$$g(0) = 1, \quad g'(0) = 0, \quad g''(0) = 0; \qquad g(a) \to 0 \text{ as } a \to +\infty; \tag{24}$$

it is convenient to use the Laplace transform $G(\Gamma)$ so that

$$g(a) = \int_0^\infty G(\Gamma)\, e^{-c\Gamma a}\, d\Gamma.$$

Dyson shows, *by using the condition* $g(a) \to 0$ *as* $a \to \infty$, that

$$S_g(t) = e^{iH_0 t/\hbar} S_g\, e^{-iH_0 t/\hbar}, \tag{25}$$

where S_g is independent of time and H_0 is the Hamiltonian of the non-interacting fields expressed in interaction representation variables.

The state function $\Phi_g(t)$ obeys the equation

$$i\hbar \frac{\partial \Phi_g(t)}{\partial t} = H'(t)\,\Phi_g(t), \tag{26}$$

where

$$H'(t) = \bar{S}_g(t)\left[H_1(t) - i\hbar \frac{d}{dt}\right] S_g(t). \tag{27}$$

Apart from a counter term due to a transient boson self energy, (27) can be written

$$H'(t) = \bar{S}_g(t)\left\{-i\hbar \frac{\Delta}{\Delta t}\right\} S_g(t), \tag{27'}$$

where the operator $\Delta/\Delta t$ acts only on $g(t)$ in (23). $H'(t)$ is the 'low frequency part' of the interaction Hamiltonian, so that in the limit of all frequencies being considered as low the intermediate representation becomes the Schrödinger representation, while in the opposite extreme case it becomes the Heisenberg representation. Because of (25) it is possible to write

$$H'(t) = e^{iH_0 t/\hbar} H'\, e^{-iH_0 t/\hbar}. \tag{27''}$$

So by (27)

$$H_0 + H' = \bar{S}_g (H_0 + H_1)\, S_g, \tag{28}$$

where we use time-independent operators on each side of the equation. The corresponding Schrödinger equation is

$$i\hbar \frac{\partial \Phi_0(t)}{\partial t} = (H_0 + H')\,\Phi_0(t), \quad \text{where} \quad \Phi_0(t) = e^{-iH_0 t/\hbar}\Phi_g(t). \tag{29}$$

The important equation (28) has been derived by using the interaction representation, but as it stands it is expressed in terms of free particle variables at one time, so that (28) is rather similar to (6) of § 1; the only difference being that S_g does not by itself ensure that H' is a V part (i.e. a sum of D-products).

Before further discussion of the transformation which is required to change H' into a V part the form of H' must be exhibited.

It follows from (27′) that

$$H'(t) = -i\hbar \sum_{n=0}^{\infty} \left(\frac{i}{\hbar}\right)^n \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t_1} dt_2 \dots \int_{-\infty}^{t_{n-1}} dt_n [H(x_1, x_n)],$$

$$[H(x_1, x_{n-1}), [\dots, [H(x, x_1), \Delta/\Delta t]] \dots] = i\hbar \Sigma I_n, \tag{30}$$

where
$$I_n = -\frac{ic}{\hbar} \int_0^{\infty} \dots \int_0^{\infty} r_1 \prod_{i=1}^{n} G(\Gamma_i)\, d\Gamma_1 \dots d\Gamma_n$$

$$\times \int_{-\infty}^{t} dt_1 \int d^3x_1 \exp[-c(\Gamma_1 + \dots + \Gamma_n)(t - t_1)]\, J(x_1; \Gamma_2, \Gamma_3, \dots, \Gamma_n), \tag{31}$$

and $J_n(x, \Gamma_1, \Gamma_2, \dots, \Gamma_n)$ is the nth order transformed matrix element of H_1 in the intermediate representation; the evaluation and the renormalization of terms of this type have been discussed in detail by Dyson (2). Defining $J_n(p; \Gamma_1, \dots, \Gamma_n)$ by

$$\frac{1}{(2\pi)^4} \int J(x, \Gamma_1, \dots, \Gamma_n)\, e^{-ipx}\, d^4x \tag{32}$$

the matrix element connecting free particle states is

$$I_n = -\frac{ic}{\hbar} \int_0^{\infty} \dots \int_0^{\infty} \Gamma_1 \prod_{i=1}^{n} G(\Gamma_i)\, d\Gamma_1 \dots d\Gamma_n\, J_{n-1}(p, \Gamma_2, \dots, \Gamma_n) \frac{e^{-ip_0 ct}}{(\Gamma_1 + \Gamma_2 + \dots + \Gamma_n) - ip_0}, \tag{33}$$

where momentum is conserved and $(\hbar c p_0)$ is the energy difference between the initial and the final states. Omitting the factor $e^{-ip_0 ct}$, replacing I_n by I'_n, and using (30) and (27″) we have

$$H' = i\hbar \sum_{n=1} I'_n. \tag{34}$$

The terms I'_n are renormalized, so H' is expressed by (34) as a series of finite matrix elements. Further the condition $g''(0) = 0$ (cf. (24)) ensures that the behaviour of H' for high frequencies is such that further manipulations with H' do not lead to further divergences.

Consider the lowest order contributions to H'. The first two terms are

$$H' = i\hbar(I'_1 + I'_2),$$

where
$$\left.\begin{aligned} H'_1 &= i\hbar I'_1 = c \int_0^{\infty} G(\Gamma)\, \Gamma\, d\Gamma \frac{H_1(p_0)}{r - ip_0}, \\ H'_2 &= i\hbar I'_2 = c \int_0^{\infty} G(\Gamma_1)\, \Gamma_1 d\Gamma_1 \int_0^{\infty} G(\Gamma_2)\, d\Gamma_2 \frac{J_1(p_0, \Gamma_2)}{\Gamma_1 + \Gamma_2 - ip_0}. \end{aligned}\right\} \tag{35}$$

$H_1(p_0)$ is that Fourier component of H_1 which produces an energy change $\hbar c p_0$ between initial and final states. The I'_1 term cannot be diagonal in the number of particles so this term has to be removed by the transformation T. Choosing Y_1 so that

$$i[H_0, Y_1]_- = -H'_1, \tag{36}$$

then to order g^2 we have

$$\tilde{T}(H_0 + H'_1 + H'_2)\, T = H_0 + \frac{i}{2}[H'_1, Y_1]_- + H'_2 \dots. \tag{37}$$

The matrix element $Y_1(p_0)$ of Y_1 which produces an energy change $\hbar c p_0$ is by (36) and (18)

$$Y_1(p_0) = \lim_{\gamma \to 0} \left(-\frac{1}{\hbar c} \frac{H'_1(p_0)}{\gamma - ip_0} \right)$$

$$= -\frac{1}{\hbar} \lim_{\gamma \to 0} \int_0^{\infty} G(\Gamma)\, \Gamma d\Gamma \frac{H_1(p_0)}{(\gamma - ip_0)(\Gamma - ip_0)}. \tag{38}$$

Thus the g^2 terms in (37) are

$$\lim_{\gamma\to 0}\int_{-\infty}^{\infty}dp_0\int_{-\infty}^{+\infty}dp_0'\int_0^{\infty}G(\Gamma_1)\,d\Gamma_1\int_0^{\infty}G(\Gamma_2)\,d\Gamma_2\,i[H_1(p_0-p_0'),H_1(p_0')]$$
$$\times\left\{\frac{\Gamma_1}{(\Gamma_1+\Gamma_2-ip_0)(\Gamma_2-ip_0+ip_0')}-\frac{\frac{1}{2}\Gamma_1\Gamma_2}{(\Gamma_1-ip_0+ip_0')(\Gamma_2-ip_0')(\gamma-ip_0')}\right\}. \tag{39}$$

Using (2) we have

$$[H_1(s_0),H_1(q_0)]_- = g^2\int d^3\mathbf{m}\,d^3\mathbf{m}'\,d^3\mathbf{l}\,d^3\mathbf{l}'\,\tilde{a}_{m'}\tilde{a}_{l'}a_m a_l V_{m'l'ml}(s_0,q_0),$$

where

$$V_{m'l'ml}(s_0,q_0)=\frac{\hbar^2}{2\epsilon_{|m-m'|}}\{\delta(E_{m'}-E_m-\epsilon_{|m-m'|}-cs_0)\,\delta(E_{l'}-E_l+\epsilon_{|m-m'|}-cq_0)$$
$$-\delta(E_{m'}-E_m+\epsilon_{|m-m'|}-cs_0)\,\delta(E_{l'}-E_l-\epsilon_{|m'-m|}-cq_0)\}.$$

From (39), if we write

$$F_{m'l'ml}=\lim_{\gamma\to 0}\int_{-\infty}^{+\infty}dp_0\int_{-\infty}^{\infty}dp_0'\int_0^{\infty}G(\Gamma_1)\,d\Gamma_1\int_0^{\infty}G(\Gamma_2)\,d\Gamma_2$$
$$\times\left\{\frac{\Gamma_1}{(\Gamma_1+\Gamma_2-ip_0)[\Gamma_2-i(p_0-p_0')]}-\frac{\frac{1}{2}\Gamma_1\Gamma_2}{[\Gamma_1-i(p_0-p_0')](\Gamma_2-ip_0')(\gamma-ip_0')}\right\}$$
$$\times iV_{m'l'ml}(p_0'-p_0',p_0), \tag{40}$$

the deuteron energy eigenvalue $\mathscr{E}$ is determined by the integral equation

$$(E_{m'}+E_{l'})\langle|\,m'l'\rangle+\Sigma F_{m'l'ml}\langle|\,ml\rangle=\mathscr{E}\langle|\,m'l'\rangle. \tag{41}$$

4. In this section we return to the problem of the existence of the T and V operators and discuss the role of unitary transformations in general.

Let us require in § 2 that a unitary operator T exists, such that $\tilde{T}HT=H_0$ (i.e. $V=0$). By the procedure outlined, this entails solving

$$i[H_0,Y_n]_-=R_n. \tag{42}$$

As remarked before this equation can always be solved except for the energy-conserving matrix elements of R_n. Thus if such energy-conserving matrix elements are explicitly separated in R, and designated as K, then there does exist a unitary operator T, such that $\tilde{T}HT=H_0+K$†.

Here then we have an apparent paradox. A unitary operator seems to transform H to a (collision) matrix (H_0+K) with a different spectrum. This, however, is not a novel situation in field theory. If in the above we do not explicitly isolate the energy-conserving matrix elements from R_n, but instead solve the equation (42) (with $V=0$), as in § 2, employing interaction representation operators, and integrating with respect to time, it can be verified step by step that

$$T=S(t_0,-\infty),$$

† Heitler has studied the K matrix ((5), § 15). He has also given the explicit form for the T matrix in this case.

where $S(t_0, -\infty)$ is the usual semi-infinite S-matrix. (Starting the other way round we can also verify this statement directly by a series expansion of

$$\tilde{S}(t_0, -\infty)\, H(t_0)\, S(t_0, -\infty)$$

(cf. Dyson (2)) when the result is found simply to be $H_0(t_0)$.) It is important that from this point of view, the $e^{-\gamma t}$ factors are introduced merely to avoid difficulties of poles connected with energy-conserving matrix elements. This type of factor has nothing to do with the behaviour at $-\infty$ or with the physical switching off of charge and the adiabatic theorem. The behaviour at $-\infty$ is a problem which never arises in the essentially 3-dimensional formulation presented in § 2.

Thus there do exist matrices T which by construction are unitary, and are such that the transformed matrix (H_0+K, or H_0) has a different spectrum from the original operator H. The only explanation seems to be that the unitary property of such matrices has been guaranteed only in terms of a series expansion, and therefore this expansion is not convergent†.

Following this argument up then, if a unitary transformation matrix T exists, such that $\tilde{T}HT = H_0 + V$ where $H_0 + V$ has the same spectrum as H, it is at least conceivable that the series defining T is convergent and T does possess double-sided unitarity (S. T. Ma (7)).

Before carrying this argument further, let is rewrite the work of § 2 in a more compact form. Let us assume V is given. Expressing V in terms of interaction representation operators $\hat{a}$, let

$$R(t_0) = 1 - (i/\hbar)\int_{-\infty}^{t_0} \hat{V}(t)\, dt + (-i/\hbar)^2 \int_{-\infty}^{t_0} dt_1 \int_{-\infty}^{t_1} dt_2\, \hat{V}(t_1)\, \hat{V}(t_2) + \ldots.$$

By the same chain of arguments which give $\tilde{S}HS = H_0$, we have

$$\tilde{R}(t_0)\,[\hat{H}_0 + \hat{V}]\, R(t_0) = H_0.$$

The bound state representation operators A are therefore related to interaction representation operators by the relation,

$$A(t_0) = \tilde{R}(t_0)\, \hat{a}(t_0)\, R(t_0).$$

This shows that

$$T(t_0) = \tilde{R}(t_0)\, S(t_0, -\infty),$$

all operators being expressed in terms of interaction representation operators.

In all the above expressions we ought properly to have isolated energy-conserving matrix elements from T and incorporated them with V explicitly. However such matrix elements form a set of measure zero in the space defined by the totality of the matrix elements V and even if separated would not contribute anything as far as bound states are concerned. (Heitler had to isolate such matrix elements because he was dealing with collision problems.)

The conclusion of this discussion is, that a criterion for the correct choice of V is that the resulting matrix R makes the power series expansion of the product $\tilde{R}S$ converge.

† The trouble may lie deeper. According to a theorem of Von Neumann (*J. reine angew. Math.* 161 (1929), 208) a *matrix* representing an unbounded operator can be transformed into any other given matrix representing a different (unbounded) operator, after a *finite* number of unitary transformations. We are indebted to Dr F. Smithies for pointing this out.

REFERENCES

(1) DANCOFF, S. M. *Phys. Rev.* 78 (1950), 382.
(2) DYSON, F. J. *Phys. Rev.* 83 (1951), 608, 1207.
(3) DYSON, F. J. *Phys. Rev.* 90 (1953), 994.
(4) GELL-MANN, M. and LOW, F. *Phys. Rev.* 84 (1951), 350.
(5) HEITLER, W. *Quantum theory of radiation*, 3rd ed. (Oxford, 1954).
(6) LEVY, M. M. *Phys. Rev.* 88 (1952), 725.
(7) MA, S. T. *Phys. Rev.* 87 (1952), 652.
(8) MATTHEWS, P. T. and SALAM, A. *Proc. roy. Soc.* A, 221 (1953), 128.
(9) SALPETER, E. E. and BETHE, H. A. *Phys. Rev.* 84 (1951), 1232.
(10) SCHWINGER, J. *Phys. Rev.* 75 (1949), 651.
(11) TAMM, I. *J. Phys., Moscow*, 9 (1945), 449.
(12) TAYLOR, J. C. *Phys. Rev.* 95 (1954), 1313.

CHRIST'S COLLEGE
CAMBRIDGE

ST JOHN'S COLLEGE
CAMBRIDGE

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 230, pp. 266–271, 1955

On the relativistic equation for scattering

BY N. KEMMER, *Department of Mathematical Physics, University of Edinburgh*
AND ABDUS SALAM, *St John's College, University of Cambridge*

(*Communicated by N. F. Mott, F.R.S.—Received* 19 *January* 1955)

By defining an analytic continuation method, Wick has been able to elucidate the structure of the relativistic equation for bound states. In particular, the equation acquires an 'elliptic' rather than a 'hyperbolic' metric. Taking advantage of a gap in the rest-mass spectrum of one-nucleon states, a similar investigation is here carried through for the relativistic equation for two-nucleon scattering, in the energy region $2\kappa < E < 2\kappa + \mu$, κ and μ being nucleon and meson masses.

1. INTRODUCTION

Wick (1954) has recently made an important advance in the study of the Bethe–Salpeter (1951) amplitude for two-nucleon bound states by an analytic continuation of the amplitude to complex values of the relative time. As a result, the integral equation (in the 'ladder' approximation) for the amplitude acquires a positive definite Euclidian metric and the mathematical discussion of its behaviour takes on a much simpler form.

It is the purpose of this paper to consider the possibility of this type of analytic continuation for the case of neutron-proton scattering. Wick's proof depends essentially on the following remark. If all real states are specified as eigenstates of operators P_i and the operator $M = +(P_0^2 - P_i^2)^{\frac{1}{2}}$, where P_0 and P_i ($i = 1, 2, 3$) are the displacement operators for energy and momentum, then for one-nucleon states there is a gap in the spectrum of the rest-mass operator M between $M = 0$ (eigenvalue for vacuum) and $M = \kappa$ (eigenvalue for one free nucleon). The investigation below started from the remark that for one-nucleon states, there is also another gap between $M = \kappa$ and $M = \kappa + \mu$, where μ is the π meson mass. Taking advantage of this gap, we consider the scattering of two nucleons in the (centre of mass system) energy range

$$2\kappa < E < 2\kappa + \mu.^*$$

We shall find that a continuation of the integral equation is indeed possible. Our results will also show the intimate connexion of the possibility of this type of continuation and the existence of creation thresholds and displaced poles (Dyson 1949).

2. EQUATION FOR THE FEYNMAN AMPLITUDE FOR SCATTERING

The Feynman amplitude (Matthews & Salam 1953) for proton-neutron scattering is defined as

$$\Psi(12) = (0 \,|\, T\psi_P(1)\,\psi_N(2) \,|\, E). \tag{1}$$

$|\, E)$ is the one-proton, one-neutron state of total energy E and zero momentum. It

* It is assumed that no bound states of the π-meson nucleon system exist. If there are such states, with lowest rest mass ρ, our considerations hold for the energy range $2\kappa < E < \kappa + \rho$, rather than $2\kappa < E < 2\kappa + \mu$.

is identified with the initial free state $|\,)_f$ at $-\infty$, by the relation (Freese 1953) (true only for scattering states),

$$|\,E) = S(0,-\infty)\,|\,\mathbf{K},-\mathbf{K})_f. \tag{2}$$

$\mathbf{K}$ and $-\mathbf{K}$ are respectively the free initial proton and neutron momenta, and $E = 2(\mathbf{K}^2+\kappa^2)^{\frac{1}{2}}$. $\Psi(12)$ satisfies an inhomogeneous integral equation which can be directly obtained by a method due to Gell-Mann & Low (1951). Using the relation

$$|\,0) = S^{-1}(\pm\infty,0)\,|\,0)_f, \tag{3}$$

which connects the true vacuum state $|\,0)$ with the free vacuum $|\,0)_f$,

$$\Psi(12) = {}_f(0\,|\,TS(\infty,-\infty)\,\psi_P(1)\,\psi_N(2)\,|\,\mathbf{K},-\mathbf{K})_f. \tag{4}$$

$\psi_P(1), \psi_N(2)$ are interaction representation operators. Equation (4) then leads to an integral equation for $\Psi(12)$. In the lowest (ladder) approximation,

$$\Psi(12) = \Psi_0(12) - \tfrac{1}{8}g^2\int S_{FP}(13)\,S_{FN}(24)\,D_F(34)\,\Psi(34)\,\mathrm{d}3\mathrm{d}4 \tag{5}$$

for neutral scalar theory.

Here $$\Psi_0(12) = {}_f(0\,|\,T\psi_P(1)\,\psi_N(2)\,|\,\mathbf{K},-\mathbf{K})_f. \tag{6}$$

It will be convenient also to introduce

$$\begin{aligned}\chi(12) &= \mathscr{D}_P(1)\,\mathscr{D}_N(2)\,\Psi(12)\\ &= g^2(0\,|\,T\psi_P(1)\,\phi(1)\,\psi_N(2)\,\phi(2)\,|\,E),\end{aligned} \tag{7}$$

where $\mathscr{D}_P(1)$, $\mathscr{D}_N(2)$ are Dirac operators. It is easy to see using (5), or directly by the procedure indicated above, that

$$\chi(12) = \tfrac{1}{2}D_F(12)\,\Psi_0(12) - \tfrac{1}{8}g^2\int D_F(12)\,S_{FP}(13)\,S_{FN}(24)\,\chi(34)\,\mathrm{d}3\mathrm{d}4. \tag{8}$$

Now for $t_1 > t_2$

$$\begin{aligned}\chi(12) = &\sum_{\mathbf{p}}(0\,|\,\psi_P(1)\,\phi(1)\,|\,\mathbf{p},\kappa)\,(\mathbf{p},\kappa\,|\,\psi_N(2)\,\phi(2)\,|\,E)\\ &+\sum_R(0\,|\,\psi_P(1)\,\phi(1)\,|\,R)\,(R\,|\,\psi_N(2)\,\phi(2)\,|\,E).\end{aligned}$$

Here $|\,\mathbf{p},\kappa)$ are the totality of one-proton states with rest mass $M=\kappa$, while $|\,R)$ are the states with other rest masses, the minimum rest mass concerned being $M = \kappa+\mu$. We shall now show that

$$(0\,|\,\psi_P(1)\,\phi(1)\,|\,\mathbf{p},\kappa) = 0. \tag{9}$$

The proof consists in remarking that the one-proton state of momentum $\mathbf{p}$ and rest mass κ is a steady state. This, besides the vacuum state, is the only Heisenberg state, for which, after a self-energy renormalization, a result similar to (3) above holds.

Thus $$(0\,|\,\psi_P(1)\,\phi(1)\,|\,\mathbf{p},\kappa) = {}_f(0\,|\,TS(\infty,-\infty)\,\psi_P(1)\,\varphi(1)\,|\,\mathbf{p},\kappa)_f. \tag{10}$$

This, on evaluation, gives $\int \mathrm{d}x_1\Sigma(x-x_1)\,{}_f(0\,|\,\psi(x_1)\,|\,\mathbf{p},\kappa)_f$, where $\Sigma(x-x_1)$ is the total contribution from self-energy graphs. After renormalizing mass, and using

$i\gamma p + \kappa = 0$, $p^2 + \kappa^2 = 0$, expression (10) indeed vanishes. It is important for the proof that the operators ψ and ϕ are defined at the same space-time point. In general $(0 \mid T\psi(1)\phi(2) \mid \mathbf{p}, \kappa)$ is not zero. Thus

$$\begin{aligned}\chi(12) &= \sum_R (0 \mid \psi_P(1)\phi(1) \mid R)(R \mid \psi_N(2)\phi(2) \mid E) \quad \text{for} \quad t_1 > t_2, \\ &= -\sum_R (0 \mid \psi_N(2)\phi(2) \mid R)(R \mid \psi_P(1)\phi(1) \mid E) \quad \text{for} \quad t_2 > t_1. \end{aligned} \tag{11}$$

Since all states $\mid R)$ have rest masses greater than $\kappa + \mu$, Wick's 'stability criteria' apply to our case as well. Writing

$$\chi(x) = \chi(12)\, e^{\frac{1}{2}iE(t_1+t_2)}, \tag{12}$$

we have, for $t > 0$, $\quad \chi(x) = \int d\mathbf{p} \int_{w_{\min.}}^{\infty} dw f(\mathbf{p}, w) \exp(i\mathbf{p}x - iwt), \tag{13}$

where

$$\left.\begin{aligned} w &= (M^2 + \mathbf{p}^2)^{\frac{1}{2}} - \tfrac{1}{2}E, \\ w_{\min.} &= [(\kappa+\mu)^2 + \mathbf{p}^2]^{\frac{1}{2}} - (\kappa + \tfrac{1}{2}\mu) > \tfrac{1}{2}\mu > 0. \end{aligned}\right\} \tag{14}$$

$\chi(x)$ therefore, for $t > 0$, is a superposition of positive frequency terms only. Similarly for $t < 0$, $\chi(x)$ contains only negative frequencies. Following Wick, we set

$$\begin{aligned}\phi_1(\mathbf{p}, p_0) &= \frac{1}{(2\pi)^4} \int d\mathbf{x}\, e^{-i\mathbf{p}\mathbf{x}} \int_0^{\infty} dt\, e^{ip_0 t} \chi(x) \\ &= \frac{1}{2\pi i} \int_{w_{\min.}}^{\infty} dw f(\mathbf{p}, w)/(w - p_0 - i\epsilon), \end{aligned} \tag{15}$$

with a corresponding definition for ϕ_2.

$\phi_1(\mathbf{p}, p_0)$ defines an analytic function of p_0 in the entire complex plane in the region $-2\pi < \arg(p_0 - w_{\min.}) \leqslant 0$, while ϕ_2 defines an analytic function in $-\pi < \arg(p_0 - w_{\max.}) < \pi$; $(w_{\max.} = -w_{\min.})$. The function $\phi(\mathbf{p}, p_0) = \phi_1 + \phi_2$, the Fourier transform of χ, is defined in the complex p_0 plane with two cuts from $w_{\min.}$ to $+\infty$ and from $-\infty$ to $w_{\max.}$. Analytic continuation from the lower to the upper half-plane is ensured through the gap between $w_{\max.}$ and $w_{\min.}$.

3. Transformation of the integral equation

Write

$$\psi_0 = u_P(\mathbf{K})\, u_N(-\mathbf{K})\, e^{i\mathbf{K}(\mathbf{x}_1 - \mathbf{x}_2)} e^{-i\frac{1}{2}E(t_1+t_2)}. \tag{16}$$

In momentum space, (8) reduces to

$$\begin{aligned}\Phi(p) &= \frac{1}{i} u(\mathbf{K})\, d(\mathbf{p} - \mathbf{K}, p_0) \\ &\quad + \frac{ig^2}{(2\pi)^4} \int d(\mathbf{p} - \mathbf{q}, p_0 - q_0)\, s_P(\mathbf{q}, q_0 + \tfrac{1}{2}E)\, s_N(-\mathbf{q}, -q_0 + \tfrac{1}{2}E)\, \Phi(q)\, d\mathbf{q}\, dq_0. \end{aligned} \tag{17}$$

Here

$$\left.\begin{aligned} d(\mathbf{k}, k_0) &= (\mathbf{k}^2 - k_0^2 + \mu^2 - i\epsilon)^{-1}, \\ s(\mathbf{k}, k_0) &= (\gamma\mathbf{k} - \gamma_0 k_0 - i\kappa - \epsilon)^{-1}, \quad \gamma_i^2 = -\gamma_0^2 = 1. \end{aligned}\right\} \tag{18}$$

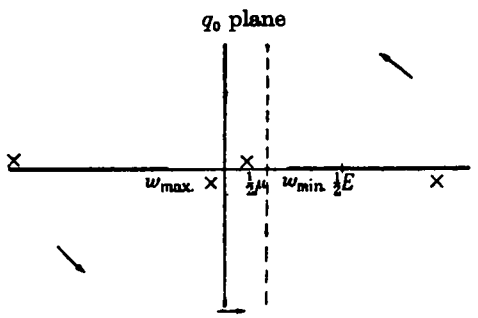
The analytic continuation of the left-hand side has been considered. On the right-hand side the first term has poles at $p_0 = \pm[(\mathbf{p} - \mathbf{K})^2 + \mu^2]^{\frac{1}{2}} \mp i\epsilon$. These poles never lie within the gap from $w_{\max.}$ to $w_{\min.}$. In fact, they always lie on the cut lines, and thus need not be considered any further.

The q_0 integration in (17) is along the real axis, the integrand having six poles. We want to rotate the contour so that it lies along an imaginary axis.

The kernel that Wick considered had just one factor $d(\mathbf{p}-\mathbf{q}, p_0-q_0)$. Taking advantage of the fact that d is a function of p_0 as well as of q_0, Wick was able to show that the poles of d cause no real difficulties in the attempt to rotate the contour in q_0 plane. In our case, besides d, there are two additional factors s_P and s_N, which are functions of q_0 alone. It is the poles of these at

$$q_0 = \tfrac{1}{2}E \pm (\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} \pm i\epsilon \quad \text{and} \quad q_0 = -\tfrac{1}{2}E \pm (\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} \pm i\epsilon$$

which have to be considered (figure 1). It is clear that if the contour is closed so as to lie along $q_0 = \frac{1}{2}\mu + iq_4$, no pole is enclosed in the first quadrant, while the pole at



× marks the positions of the poles.

FIGURE 1

$q_0' \equiv -\frac{1}{2}E + (\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} - i\epsilon$ is enclosed only for $\mathrm{Re}\,(q_0') < \frac{1}{2}\mu$. The residue R, at this pole, gives

$$R = \frac{-g^2}{16\pi^3}\int d(\mathbf{p}-\mathbf{q}, p_0-\tfrac{1}{2}E-\sqrt{(\mathbf{q}^2+\kappa^2)})\, s_P(\mathbf{q}, \sqrt{(\mathbf{q}^2+\kappa^2)}) \times \frac{\boldsymbol{\gamma}\mathbf{q}-\gamma_0(\sqrt{(\mathbf{q}^2+\kappa^2)}-\frac{1}{2}E)}{(\sqrt{(q^2+\kappa^2)}-E-i\epsilon)}\,\Phi(\mathbf{q}, \sqrt{(\mathbf{q}^2+\kappa^2)}-\tfrac{1}{2}E)\,d\mathbf{q}, \qquad (19)$$

the integrating being over the volume of the finite sphere

$$\sqrt{(\mathbf{q}^2+\kappa^2)} < \tfrac{1}{2}(E+\mu).$$

The integrand in (19) has the expected behaviour, at $(\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} = \frac{1}{2}E$, corresponding to the coincidence of poles in (17). For $q_0 = 0$, $(\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} = \frac{1}{2}E$, real nucleon emission and rescattering is possible and the singularity of the integrand, with its behaviour as a δ function plus a principal part integration, correctly describes just this possibility. A detailed discussion of this has been given by Eden (1950) and Hamilton (1952).* The result of this rotation then is

$$\Phi(\mathbf{p}, p_0) = \frac{1}{i}u(\mathbf{K})\,d(\mathbf{p}-\mathbf{K}, p_0) - \frac{g^2}{(2\pi)^4}P\int d'(\mathbf{p}-\mathbf{q}, q_4-\tfrac{1}{2}i\mu+ip_0) \times s_P'\{\mathbf{q}, q_4-\tfrac{1}{2}i(\mu+E)\}\, s_N'\{-\mathbf{q}, -q_4+\tfrac{1}{2}i(\mu-E)\}\,\Phi(\mathbf{q}, iq_4+\tfrac{1}{2}\mu)\,d\mathbf{q}\,dq_4 + R. \qquad (20)$$

Here the q_4 integration goes from $-\infty$ to $+\infty$;

$$d'(\mathbf{q}, q_4) = (\mathbf{q}^2+q_4^2+\mu^2)^{-1}, \quad s'(\mathbf{q}, q_4) = (\boldsymbol{\gamma}\mathbf{q}+\gamma_4 q_4 - i\kappa)^{-1}, \qquad (21)$$

* If we had closed the contour along the imaginary axis itself, the poles in (17) would always coincide on the path of integration. Our choice of the ordinate $q_0 = \frac{1}{2}\mu + iq_4$, makes them coincide at an interior point of the sphere $(\mathbf{q}^2+\kappa^2)^{\frac{1}{2}} < \frac{1}{2}(E+\mu)$.

and $\gamma_4 = -i\gamma_0$. One of the denominators in the integrand,

$$[\mathbf{q}^2 + \{q_4 - \tfrac{1}{2}i(\mu + E)\}^2 + \kappa^2]$$

can vanish for the two-dimensional surface $q_4 = 0$, $\mathbf{q}^2 + \kappa^2 = \frac{1}{4}(E+\mu)^2$, and the principal value P refers to this possibility. The integration in R is a three-dimensional integration extending through the region $\mathbf{q}^2 + \kappa^2 < \frac{1}{4}(E+\mu)^2$.

Finally, an analytic continuation in the p_0 plane can now be carried through. We choose to continue from real $p_0 > \frac{1}{2}\mu$ to $p_0 = ip_4 + \frac{1}{2}\mu\,(p_4 > 0)$; and for $p_0 < \frac{1}{2}\mu$ to $p_0 = \frac{1}{2}\mu + ip_4\,(p_4 < 0)$. Defining $\Psi'(\mathbf{p}, p_4) = i\Phi(\mathbf{p}, \frac{1}{2}\mu + ip_4)$, we obtain

$$\begin{aligned}\Psi'(\mathbf{p}, p_4) = u(\mathbf{K})\, d'(\mathbf{p} - \mathbf{K}, p_4 - \tfrac{1}{2}i\mu) - \frac{g^2}{(2\pi)^4} P \int d'(\mathbf{p} - \mathbf{q}, p_4 - q_4) \\ \times s'_P\{\mathbf{q}, q_4 - \tfrac{1}{2}i(\mu + E)\}\, s'_N\{-\mathbf{q}, -q_4 + \tfrac{1}{2}i(\mu - E)\}\, \Psi'(\mathbf{q}, q_4)\, d\mathbf{q}\, dq_4 \\ + iR\,(\text{with its argument } p_0 \text{ changed to } ip_4 + \tfrac{1}{2}\mu). \qquad (22)\end{aligned}$$

We have thus succeeded in obtaining an elliptic metric. As already stated, in the second term on the right there is a two-dimensional surface (in four-dimensional space) where one of the denominators in the integrand vanishes. The real price, however, has been paid in the appearance of the term R with its three-dimensional integration over a finite sphere.

Mixed integral equations of the type (22) are in principle exactly soluble (see Appendix). In practice, however, one would set up an iteration between the second and the third terms on the right-hand side. Thus, as a first step, the inhomogeneous term, ud', is substituted for Ψ' in R, and the three-dimensional integration performed. The sum of ud' and R now serves as a new inhomogeneous term for the four-dimensional integral equation. Physically, this means that we are treating all virtual processes properly, but are limiting ourselves to the case when a formation and rescattering of real nucleons is allowed only once in the interaction zone.

To calculate the scattering cross-section, we require $\Phi(\mathbf{p}, p_0)$ for $\mathbf{p} = \mathbf{K}'$ and $p_0 = 0$, where $\mathbf{K}'$, $-\mathbf{K}'$ are the final nucleon momenta. In other words we need $\Psi'(\mathbf{K}, \frac{1}{2}i\mu)$. On account of the analytic properties of Ψ' however, it would suffice to solve (22) for a real argument p_4 and to continue to the complex value $p_4 = \frac{1}{2}i\mu$. If (22) is solved numerically, this can be achieved as follows. If in the q_0 plane we rotate the contour on to $q_0 = -\frac{1}{2}\mu + iq_4$ (rather than $q_0 = \frac{1}{2}\mu + iq_4$) we obtain an equation analogous to (22) for $\Phi(\mathbf{p}, -\frac{1}{2}\mu + ip_4)$, which need not be written down. Setting

$$\Psi''(\mathbf{p}, p_4) = i\Phi(\mathbf{p}, -\tfrac{1}{2}\mu + ip_4), \qquad (23)$$

this equation for Ψ'' also can be solved for a real argument p_4. Since $\Phi(\mathbf{K}', p_0)$ is an analytic function of p_0 in the strip enclosed by the ordinates $-\frac{1}{2}\mu + ip_4$ and $\frac{1}{2}\mu + ip_4$

$$\Phi(\mathbf{K}', 0) = \frac{1}{2\pi i} \int_C \frac{\Phi(\mathbf{K}', p_0)}{p_0}\, dp_0, \qquad (24)$$

where C is the closed contour along these ordinates. Thus

$$\begin{aligned}\Phi(\mathbf{K}', 0) &= \frac{1}{2\pi i} \int \frac{\Phi(\mathbf{K}', \frac{1}{2}\mu + ip_4)}{p_4 - \frac{1}{2}i\mu} - \frac{\Phi(\mathbf{K}', -\frac{1}{2}\mu + ip_4)}{p_4 + \frac{1}{2}i\mu}\, dp_4 \\ &= \frac{1}{2\pi} \int \left[\frac{\Psi'(\mathbf{K}', p_4)}{p_4 - \frac{1}{2}i\mu} - \frac{\Psi''(\mathbf{K}', p_4)}{p_4 + \frac{1}{2}i\mu}\right] dp_4. \qquad (25)\end{aligned}$$

The procedure described in this paper works equally well for the meson-nucleon case, in the energy range $\kappa+\mu<E<\kappa+2\mu$. The resulting equation does not suffer from the renormalization difficulties of the three-dimensional Tamm–Dancoff methods. In the end, it may be pointed out, that unlike the case considered in the preceding paragraph (calculation of the scattering cross-section), there are a number of problems where the amplitude $\chi(x)$ is needed for all values of the relative time. One such case for example, is the relativistic treatment of the deutron photo-disintegration problem.

The authors are indebted to Mr J. C. Taylor and Professor R. E. Peierls, F.R.S., for many helpful suggestions.

Appendix

Integral equation (22) is a mixed equation of the type

$$F(x_1,x_2)=f(x_1,x_2)+\lambda\int K_1(x_1,x_2;\,y_1,y_2)\,F(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 + \lambda\int K_2(x_1,x_2;\,y_1,g(y_1))\,F(y_1,g(y_1))\,\mathrm{d}y_1,$$

or symbolically $$F=f+\lambda K_1F+\lambda\mathscr{K}_2\mathscr{F}.$$

The script type of $\mathscr{K}_2$ and $\mathscr{F}$ emphasizes the partial nature of the integral operation. Such equations can be solved exactly by methods used by Salam (1953). Assume that the Fredholm determinant $d_1(\lambda)$ for the kernel λK_1 exists. Then

$$F=\frac{1}{1-\lambda K_1}\frac{\mathrm{d}_1(\lambda)}{\mathrm{d}_1(\lambda)}[f+\lambda\mathscr{K}_2\mathscr{F}]=f'+\mathscr{K}'\mathscr{F}.$$

Setting $x_2=g(x_1)$ in F, f' and $\mathscr{K}'$, we immediately get an integral equation for $\mathscr{F}$, which is solved by conventional methods.

The kernel in the integral equation (22) is singular, in the sense that all its traces diverge logarithmically. A modification of $\frac{1}{q^2+\mu^2}$ to $\frac{1}{q^2+\mu^2}-\frac{1}{q^2+M^2}$ removes this difficulty. Here M is taken to be arbitrarily large. This type of Feynman cut-off seems justified for scattering problems, where no problems of the existence of the solution are involved (Goldstein 1953).

References

Dyson, F. J. 1949 *Phys. Rev.* **75**, 1736.
Eden, R. J. 1950 *Proc. Roy. Soc.* A, **210**, 388.
Freese, E. 1953 *Z. Naturf.* **12**, 776.
Gell-Mann, M. & Low, F. 1951 *Phys. Rev.* **84**, 350.
Goldstein, J. S. 1953 *Phys. Rev.* **91**, 1516.
Hamilton, J. 1952 *Proc. Camb. Phil. Soc.* **48**, 640.
Matthews, P. T. & Salam, A. 1953 *Proc. Roy. Soc.* A, **221**, 128.
Salam, A. 1953 *Proc. Camb. Phil. Soc.* **49**, 638.
Salpeter, E. & Bethe, H. 1951 *Phys. Rev.* **84**, 1232.
Wick, G. C. 1954 *Phys. Rev.* **96**, 1124.

A. SALAM
1956, Febbraio
Il Nuovo Cimento
Serie X, Vol. 3, pag. 424-429
© 1956 Società Italiana di Fisica

On Generalised Dispersion Relations.

A. SALAM

St. John's College - Cambridge, England

(ricevuto il 15 Dicembre 1955)

Summary. — A relation is derived which connects the real part of the non forward meson-nucleon scattering amplitude, to its imaginary part.

1. – A number of investigations [1] have recently appeared in which relations have been obtained connecting essentially the real part of the forward scattering amplitude to its imaginary part. More recent investigations are those of KARPLUS and RUDERMAN [2] and GOLDBERGER [3] who have shown the great utility of such relations for the problem of meson-nucleon scattering.

Attempts to obtain more general relations, connecting for example the real part of the scattering amplitude at an arbitrary angle in the centre of mass system, to its imaginary part at the same angle, have in general not succeeded. In this paper we shall show that relations of the required type do exist for the meson-nucleon system, though indeed they do not connect the real and the imaginary parts of the amplitude at precisely the same angle in the centre of mass system. For the proof we need the concept of micro-causality, as interpreted in ref. [1] viz: the commutator of the field operators at two points vanishes if these points have a space-like separation. Our main result is eq. (27) of this paper.

2. – Consider a field theory describing neutral mesons of mass μ, interacting with nucleons of mass $\varkappa$. The interaction Lagrangian is $j(x)\varphi(x)$, while

---

[1] GELL-MANN, GOLDBERGER and THIRRING: *Phys. Rev.*, **95**, 1612 (1954).
[2] R. KARPLUS and M. RUDERMAN: *Phys. Rev.*, **98**, 771 (1955).
[3] M. L. GOLDBERGER: *Phys. Rev.*, **99**. 979 (1955).

the causality condition demands that (4)

$$[j(x), j(y)] = 0 \qquad \text{when} \quad (x-y)^2 < 0 \tag{1}$$

Let k, k' be the initial and the final meson and p, p' (5) the corresponding nucleon 4-momenta. The Feynman matrix element for scattering is given by

$$M_F(k, k'; p, p') = i \int \mathrm{d}^4x \, \mathrm{d}^4y \exp[ik'x - iky](p' | [j(x) j(y)]_+ | p); \tag{2}$$

GOLDBERGER (3) has shown that the retarded matrix element

$$M(k, k'; p, p') = i \int \mathrm{d}^4x \, \mathrm{d}^4y \exp[ik'x - iky] \theta(x-y)(p' | [j(x), j(y)] | p) \tag{3}$$

coincides with M_F for $k_0 > 0$ and contains precisely the same information. Define

$$M_R(k; p, p') = i \int \mathrm{d}^4x \, \theta(-x)(p' | [j(0), j(x)] | p) \exp[-ikx] \tag{4}$$

$$M_A(k; p, p') = i \int \mathrm{d}^4x \, \theta(+x)(p' | [j(0), j(x)] | p) \exp[-ikx] \tag{5}$$

$$D(k; p, p') = \tfrac{1}{2}(M_R - M_A) \tag{6}$$

$$A(k; p, p') = \frac{1}{2i}(M_R + M_A) . \tag{7}$$

If M_R, M_A, D and A are considered as functions of k, p and p', the energy shell specifies the following conditions on these three vectors:

$$\text{(i)} \quad k^2 = \mu^2 \, p^2 = p'^2 = \varkappa^2 , \tag{8}$$

$$\text{(ii)} \quad p_0, \; p_0' > 0 ; \quad k_0, \; p_0 + k_0 - p_0' > 0 . \tag{9}$$

$$\text{(iii)} \quad k \cdot (p - p') = p \cdot p' - \varkappa^2 . \tag{10}$$

We now note from the definitions, that

$$M^*_{R,A}(k; p, p') = M_{R,A}(-k; p', p) , \tag{11}$$

$$\delta^4(k + p - p' - k') \, M_R(k; p, p') = -\delta^4(k + p - p' - k') \, M_A(-k'; p, p') . \tag{12}$$

---

(4) We use the metric $x^2 = x_0^2 - \boldsymbol{x}^2$.

(5) In this paper we consider only the no spin-slip case, so that spin suffixes for the nucleons will be omitted.

The last relation holds since both sides equal $(2\pi)^{-4} M(k, k'; p, p')$ (6).

Further

$$k\cdot(p+p') = k'\cdot(p+p'), \quad \text{since } p^2 = p'^2 = \varkappa^2, \tag{13}$$

$$k\cdot(p-p') = k'\cdot(p'-p), \quad \text{since } k^2 = k'^2 = \mu^2. \tag{14}$$

Write

$$M_{B,A} = \bar{u}(p')\{L'_{B,A} + i\gamma k M'_{B,A}\}u(p) ;$$

L' and M' depend invariantly only on $k\cdot(p+p')$, $k\cdot(p-p')$, $p\cdot p'$. Thus

$$M'_{B,A}(k; p, p') = M'_{B,A}(k'; p', p) \text{ etc.} \tag{15}$$

Combining (11), (12) and (15),

$$\delta^4(k+p-k'-p')(M'_B(k; p, p') + M'_A(k; p, p')) = 0 \text{ etc.} \tag{16}$$

Write $D = \bar{u}\{L'_D + i\gamma k M'_D\}u$. From (6), (15) and (16), L'_D and M'_D are real. In the next section we choose a frame, where $\bar{u}u$, $\bar{u}i\gamma ku$ are also real. So we prove the important result that both $D(k; p, p')$ and $A(k; p, p')$ are real (7) on the energy shell and are respectively the real and the imaginary parts of $M_B(k; p, p')$. It is crucial for the above proof that relations (13) and (14) should hold and this is only true if the same particles participate in the initial and the final states.

3. – The « natural » frame of reference appears to be the frame in which the nucleons are at rest. In this frame, without loss of generality, choose

$$p = (P^2+\varkappa^2)^{\frac{1}{2}}, \quad 0, \quad 0, \quad P, \tag{17}$$

$$p' = (P^2+\varkappa^2)^{\frac{1}{2}}, \quad 0, \quad 0, \quad -P. \tag{18}$$

Since $p\cdot p' - \varkappa^2 = 2P^2$, we obtain but one condition on the vector, k from the energy-shell relations; in fact

$$k = k_0, \quad k_1, \quad k_2, \quad -P. \tag{19}$$

---

(6) This is proved by using $j(x) = \exp[iPx]j(0)\exp[-iPx]$.

(7) The author is indebted to Mr. W. GILBERT who first pointed this result to him.

Also

$$\bar{u}u = p_0/k, \qquad \bar{u}i\gamma ku = k_0 .$$

Thus M_B really depends only on k_0 and P.

Considering the integral expression for M_B (eq. (4)), it is also clear that its k_0 dependence arises quite trivially from the factor $\exp[-ikx]$. This is because $(p'|[j(0), j(x)]|p)$ depends only on P in our frame of reference, through the invariants $(p+p')x$, etc. If we hold P fixed (or invariantly $p\cdot p'$ ([8]) fixed) the mathematical problem reduces to the same as presents itself for the forward scattering case and precisely the same steps as GOLDBERGER ([3]), can be followed to obtain generalized dispersion relations.
Write

$$J_1(P, x) = (p|[j(0), j(x)]|p') - (p'|[j(0), j(x)]|p) \tag{20}$$

$$J_2(P, x) = i\{(p|[j(0), j(x)]|p') + (p'|[j(0), j(x)]|p)\}. \tag{21}$$

From $M'_B(k; p, p') = M'_B(k'; p', p)$ and the evenness of $\theta(x)$ $(p'|[j(0), j(x)]|p)$ in x_1 and x_2 in the new frame, $(x = x_0, x_1, x_2, x_3)$,

$$M_B(k_0, \boldsymbol{k}; p, p') = M_B(k_0, -\boldsymbol{k}; p', p) . \tag{22}$$

Thus

$$M_B = \tfrac{1}{2}[M_B(k_0, \boldsymbol{k}; p, p') + M_B(k_0, -\boldsymbol{k}; p, p')] = \tag{23}$$
$$= \tfrac{1}{2}\int\theta(x)\exp[-ik_0x_0]\{J_1(P, x)\sin \boldsymbol{kx} + J_2(P, x)\cos \boldsymbol{kx}\}\,\mathrm{d}^4x$$

$$D(k_0, P) = \tfrac{1}{2}\int\theta(x)\cos k_0x_0\{J_1(P, x)\sin Px_3 + J_2(P, x)\cos Px_3\}\cdot \tag{24}$$
$$\cdot\cos(k_1x_1 + k_2x_2)\,\mathrm{d}^4x$$

$$A(k_0, P) = -\tfrac{1}{2}\int\theta(x)\sin k_0x_0\{J_1(P, x)\sin Px_3 + J_2(P, x)\cos Px_3\}\cdot \tag{25}$$
$$\cdot\cos(k_1x_1 + k_1x_1)\,\mathrm{d}^4x .$$

Using (24) and (25) we now verify directly that

$$\frac{D(\alpha, P) - D(\beta, P)}{\alpha^2 - \beta^2} = +\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{k_0A(k_0, P)\,\mathrm{d}k_0}{(k_0^2 - \alpha^2)(k_0^2 - \beta^2)} \tag{26}$$

([8]) In the centre of mass system, $p\cdot p' - \varkappa^2 = 2a^2\sin^2\theta/2$ where a is the magnitude of the 3-momentum in the problem.

$$(27)\qquad \frac{D(\alpha, P)-D(\beta, P)}{\alpha^2-\beta^2} = +\frac{2}{\pi} P\int_0^\infty \frac{k_0 A(k_0, P)\,\mathrm{d}k_0}{(k_0^2-\alpha^2)(k_0^2-\beta^2)}\,.$$

The proof follows by interchanging the order of x and k_0 integrations and performing a simple contour integration in the k_0 plane. Here, for the first time, the causality relation (1) is used. This part of the work is subject to the same limitations as noted by GOLDBERGER [3].

To see significance of the relation (27), consider the case of scalar nucleons. Here the k_0 dependence of D and A arises only from $k\cdot(p+p')$. In the laboratory system $p=\varkappa$, 0 0 0, $k\cdot(p+p') = 2k_0 + \varkappa(p_0'-\varkappa)$ and $p\cdot p' = \varkappa p_0'$. A variation of $k\cdot(p+p')$ while $p\cdot p'$ is held fixed is the same as the variation of the incoming meson energy k_0, while p_0', the final nucleon energy is fixed. Thus in the laboratory system, for scalar nucleons,

$$(28)\qquad \frac{\mathrm{Re}\,M(k_0^1, p_0') - \mathrm{Re}\,M(k_0^2, p_0')}{(k_0^1-k_0^2)(k_0^1+k_0^2+p_0'-\varkappa)} = $$

$$= +\frac{P}{\pi}\int \frac{(2k_0'+p_0'-\varkappa)\,\mathrm{Im}\,M(k_0, p_0')\,\mathrm{d}k_0}{(k_0-k_0^1)(k_0-k_0^2)(k_0+k_0^1+p_0'-\varkappa)(k_0+k_0^2+p_0^1-\varkappa)}\,.$$

4. – Returning to eq. (27), the imaginary part of the amplitude can, in principle, be determined from experiment for $k_0 \geqslant (\mu^2+P^2)^{\frac{1}{2}}$. In order to use relation (27) we have to determine the correct analytic continuation of $A(k_0, P)$ in the (unphysical) range $0 \leqslant k_0 < (\mu^2+P^2)^{\frac{1}{2}}$. From eq. (7),

$$(29)\qquad A(k) = \tfrac{1}{4}\int (p'|[j(0), j(x)]|p)\,\exp[-ikx]\,\mathrm{d}^4x$$

$$(30)\qquad = \frac{(2\pi)^4}{2}\sum_n (p'|j(0)|n)(n|j(0)|p)\,\delta^4(n-p-k) -$$

$$- \frac{(2\pi)^4}{2}\sum_n (p'|j(0)|n)(n|j(0)|p)\,\delta^4(n-p'+k)\,.$$

In the second term in (30), $n = p'-k$, and therefore the masses of the (physical) intermediate state are given by $(\varkappa^2+\mu^2-2p'\cdot k)^{\frac{1}{2}}$. The only possible state which can contribute is the single nucleon state corresponding to $\mu^2-2p'\cdot k = 0$. Its contribution to $A(k_0, P)$ for the no spin-flip case, with renormalized g properly defined, is

$$(31)\qquad -\frac{g^2}{4\pi}(4\pi^2 k_0)\delta(2p_0k_0-\mu^2-2P^2)\,. \quad (^9)$$

(9) Considerations on this and the spin-flip case will be given in a second paper.

Now consider the first term in eq. (30). This term does not contribute at all for the forward scattering case. In the present case, however, for k_0 in the range $(\varkappa\mu - P^2)/(P^2+\varkappa^2)^{\frac{1}{2}}$, $(\mu^2+P^2)^{\frac{1}{2}}$, we find $n^2 = (k+p)^2 \geqslant (\varkappa+\mu)^2$, and meson-nucleon states may make a contribution. Indeed this is the price one really pays for a generalized dispersion relation, and it has not been possible to find a satisfactory method for the evaluation of this contribution. It is of course possible to make a perturbation approximation to $A(k_0, P)$ just for this range.

A relation similar to eq. (27) has previously been derived by BETHE and ROHRLICH ([10]) and applied to the case of small angle scattering of light by Coulomb field. The quantity held fixed in their case is b, an impact parameter.

* * *

The author is indebted to Dr. F. SMITHIES, Mr. W. GILBERT, Professor H. A. BETHE, and Dr. P. MARTIN for discussions.

([10]) H. A. BETHE and F. ROHRLICH: *Phys. Rev.*, **86**, 10 (1952).

RIASSUNTO (*)

Si deriva una relazione che connette la parte reale dell'ampiezza dello scattering mesone-nucleone non diretto in avanti alla sua parte immaginaria.

(*) *Traduzione a cura della Redazione.*

ABDUS SALAM and W. GILBERT
1956, Marzo
Il Nuovo Cimento
Serie X, Vol. 3, pag. 607-611
© 1956 Società Italiana di Fisica

On Generalised Dispersion Relations II.

ABDUS SALAM
St. John's College - Cambridge

W. GILBERT
Trinity College - Cambridge

(ricevuto il 16 Gennaio 1956)

Summary. — Considerations of an earlier paper are generalised to write down non-forward scattering, spin-flip dispersion relations. The contribution from the « bound state » is discussed.

---

1. – This paper continues the discussion of an earlier paper [1]. A meson of 4-momentum k ($k^2 = \mu^2$) is scattered by a nucleon of initial 4-momentum p, ($p^2 = \varkappa^2$) spin λ; no other particles are emitted in the finl state, the meson and the nucleon momenta being k' and p', and spin λ'. From energy conservation,

$$k' = p + k - p', \qquad k\cdot(p - p') = p\cdot p' - \varkappa^2 . \tag{1}$$

In the earlier paper the scattering amplitude M_R defined covariantly, was written as

$$M_R(\lambda'\lambda) = \bar{u}^{\lambda'}(p')\{L' + i\gamma k M'\}u^{\lambda}(p) , \tag{2}$$

where L' and M' depend on $k\cdot(p+p')$ and $p\cdot p'$.

To obtain dispersion relations it was found necessary to work in a special

---

[1] A. SALAM: *Nuovo Cimento*, **3**, 424 (1956).

[2] Notation: $p = \boldsymbol{p}$, $p_4 = ip_0$, $\gamma_i = \begin{pmatrix} & \sigma_i \\ \sigma_i & \end{pmatrix}$, $\gamma_4 = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$. Define $i\gamma p = i\boldsymbol{\gamma}\cdot\boldsymbol{p} + i\gamma_4 p_4$; $p\cdot q = p_0 q_0 - \boldsymbol{p}\cdot\boldsymbol{q}$. Thus $(i\gamma p)^2 = p^2$.

frame of reference defined by $p+p'=0$. If $p=0,0,P$, $k=Q,0,-P$ $p\cdot p'-\varkappa^2=2P^2$, and $k\cdot(p+p')=2k_0p_0$. Further (3) in this frame

$$(3)\qquad \begin{cases} u^1(p')\{1, i\gamma k\}u^1(p) = \dfrac{p_0}{\varkappa}, k_0 \\ u^2(p')\{1, i\gamma k\}u^1(p) = 0, \quad \dfrac{PQ}{\varkappa}. \end{cases}$$

In reference (1), a dispersion relation was obtained for $M_B(11) = (p_0/\varkappa)L'+k_0M'$. It is in fact equally simple to obtain relations for L' and M' separately.

Consider $M_B(21)$, the spin-flip amplitude in this frame. $M_B(21)=-M_B(12)=(PQ/\varkappa)M'$. Now

$$(4)\qquad (p'2\,|[j(0), j(x_0, -x_1, x_2, x_3)]\,|p1) = (p'1\,|[j(0)j(x)]\,|p2)$$

$$(5)\qquad (p'2\,|[j(0), j(x_0, x_1, x_2, -x_3)]\,|p1) = -(p2\,|[j(0)j(x)]\,|p'2)\,.$$

These results follow by using the parity operator; for example

$$P_{x_3}j(x)P_{x_3}^{-1} = -j(x_0, x_1, x_2, -x_3)\,; \qquad P_{x_3}|p_0, p, 1\rangle = i|p_0, -p, 1\rangle\,.$$

From $M_B(12) = -M_B(21)$, and using (4) and (5),

$$M_B(21) = -\tfrac{1}{2}\int\theta(-x)\exp[-ik_0x_0]\sin Qx_1\{J_1(P, x)\sin Px_3 + J_2(P, x)\cos Px_3\}\,\mathrm{d}^4x.$$

Here

$$J_1(x) = \{(p'2\,|p1) - (p1\,|p'2)\} - \{(p'1\,|p2) - (p2\,|p'1)\}$$

$$J_2(x) = i\{(p'2\,|p1) + (p1\,|p'2)\} - i\{(p'1\,|p2) + (p2\,|p'1)\}$$

in an obvious notation. J_1, J_2 are real. From the contour (4) integral

$$\int\frac{\mathrm{d}k_0'\exp[-ik_0'x_0]}{k_0'^2-k_0^2}\left(\frac{\sin Q'x_1}{Q'}\right),$$

(3) The (positive energy) free particle spinors are defined as

$$u_\varkappa^\lambda(p) = \frac{1}{N(p)}(i\gamma p - k)_{\alpha\lambda}\,, \qquad \lambda = 1, 2\,; \qquad \varkappa = 1, 2, 3, 4\,.$$

We obtain $N(p)$, the « invariant » normalization from $u^{\lambda'}(p')u^\lambda(p) = \delta_{\lambda'\lambda}$. Thus $N^{-2}(p) = 2\varkappa(\varkappa + p_0)$.

(4) M. L. GOLDBERGER: *Phys. Rev.*, **99**, 979 (1955).

we infer

$$\frac{\mathrm{Re}\, M'(k_0, P)}{k_0} = \frac{2}{\pi}\,\mathrm{p.v.}\int_0^\infty \frac{\mathrm{Im}\, M'(k_0', P)\,\mathrm{d}k_0'}{k_0'^2 - k_0^2}\,. \tag{6}$$

Clearly $k_0 M'(k_0, P)$ would satisfy a dispersion relation as in eq. (27) of reference [1]. Thus also

$$\mathrm{Re}\, L'(k_0, P) = \frac{2}{\pi}\,\mathrm{p.v.}\int_0^\infty \frac{\mathrm{Im}\, L'(k_0', P) k_0'\,\mathrm{d}k_0'}{k_0'^2 - k_0^2}\,. \tag{7}$$

L_k' and M_k' are functions of invariant scalar products $k\cdot(p+p')$ and $p\cdot p'$. Relations (6) and (7) can thus be expressed entirely in terms of invariant quantities.

2. – It is perhaps useful to set down the transformation formulae from the frame used, to the conventional centre of mass frame, defined from the requirement $k^c + p^c = 0$. The superscript « c » distinguishes c.m. quantities.

Let $p_\mu^c = a_{\mu\nu} p_\nu$. The matrix $(a_{\mu\nu})$ is given by

$$a_{11} = a_{44} = \cosh\chi\,, \quad a_{14} = -a_{41} = i\sinh\chi\,, \quad a_{22} = a_{33} = 1$$

where $\mathrm{tg}\; h\chi = Q/(p_0 + k_0)$.

Define

$$S\gamma_\nu S^{-1} = a_{\mu\nu}\gamma_\mu\,.$$

For the $a_{\mu\nu}$ in (4),

$$S = \exp\left[(i/2)\,\gamma_1\gamma_4\chi\right] = c + i\gamma_1\gamma_4 s\,.$$

Here

$$c = \cosh\chi/2\,, \qquad s = \sinh\chi/2\,.$$

If $u^{(s\lambda)}(p^c) = S u^\lambda(p)$, clearly

$$\bar{u}^{\lambda'}(p')\{1, i\gamma k\}u^\lambda(p) = u^{(s\lambda')}(p'^c)\{1, i\gamma k^c\}u^{(s\lambda)}(p^c)\,. \tag{8}$$

Now

$$S u^\lambda(p)|_\alpha = \frac{1}{N(p)}\, S(i\gamma p - \varkappa)|_{\alpha\lambda} = \frac{1}{N(p)}\,(i\gamma p^c - \varkappa)S|_{\alpha\lambda}\,. \tag{9}$$

Considered as a 4×4 matrix, only two of the columns of $(i\gamma p^c - \varkappa)$ are

linearly independent. Therefore

$$Su^1(p) = [(p_0 + \varkappa)(p_0^c + \varkappa)]^{-\frac{1}{2}}\{c(p_0 + \varkappa)u^1(p^c) - Ps\,u^2(p^c)\} \tag{10}$$

$$Su^2(p) = [(p_0 + \varkappa)(p_0^c + \varkappa)]^{-\frac{1}{2}}\{Ps\,u'(p^c) + c(p_0 + \varkappa)\,u^2(p^c)\}. \tag{11}$$

Define

$$M_R^c(11) = u^1(p^c)\{L_R' + i\gamma k^c M_R'\}\,u'(p^c), \quad \text{etc.} \tag{12}$$

From (10) and (11),

$$M(11) = \alpha M^c(11) + \beta M^c(21) \ (^5) \tag{13}$$

$$M(21) = -\beta M^c(11) + \alpha M^c(21) \tag{14}$$

where

$$\beta = \frac{P \sinh \chi}{p_0^c + \varkappa}, \qquad \alpha^2 + \beta^2 = 1.$$

$M^c(11)$ corresponds to no spin-flip, and $M^c(21)$ to the spin-flip amplitudes in the c.m. system (the latter with no azimuth change). In terms of the phase shifts, therefore

$$M^c(11) = \left(\frac{4\pi}{\varkappa}\right)\frac{p_0^c + k_0^c}{k_c}\sum (la_{l-} + (l+1)a_{l+})P_l(\cos\theta_c), \tag{15}$$

$$M^c(21) = \left(\frac{4\pi}{\varkappa}\right)\frac{p_0^c + k_0^c}{k_c}\sum (a_{l-} - a_{l+})\sin\theta_c \frac{\mathrm{d}}{\mathrm{d}(\cos\theta_c)}P_l(\cos\theta_c). \tag{16}$$

Here $a = e^{2i\delta} - 1/2i$, and δ_{l-} and δ_{l+} refer to phase shifts, corresponding to $j = l - \frac{1}{2}$, and $j = l + \frac{1}{2}$. k_c and θ_c are the 3-momentum and the angle in the c.m. system. To obtain k_c and θ_c, note $p\cdot p' - \varkappa^2 = k_c^2(1 - \cos\theta_c) = 2P^2$, and $k\cdot p = k_0 p_0 + P^2 = k^c p_0^c + k_c^2$.

3. – The case of charged mesons has been completely solved by GOLDBERGER (4). Write

$$M_{\alpha\beta} = \delta_{\alpha\beta}\bar{u}(L' + i\gamma k M')u + \tfrac{1}{2}[\tau_\alpha, \tau_\beta]\bar{u}(L'' + i\gamma k M'')u.$$

---

(5) On account of our choice of axes ($p = 0,0,P$; $k = Q, 0, -P$), $M(11) = M(22)$ and $M(12) = -M(21)$. The same holds for the trasformed quantities M^c.

Then L' and M'' satisfy a relation of the type (6); M', L'' a relation of the type (7). L' and M' refer to the phase shift combination $\frac{1}{3}a(\frac{1}{2}) + \frac{2}{3}a(\frac{3}{2})$, and L'' and M'' to $a(\frac{1}{2}) - a(\frac{3}{2})$.

In relations (6) and (7), $k_0 \geqslant (\mu^2+P^2)^{\frac{1}{2}}$ is the physical range; $(\varkappa\mu - P^2)/2p_0 \leqslant k_0 < (\mu^2+P^2)^{\frac{1}{2}}$ is the « unphysical range ». Below $(\varkappa\mu - P^2)/2p_0$, the so called « bound state » contributes to Im M_R. The exact energy dependence of this term can be computed if we adopt the convention,

$$\Gamma_5(p_1^2 = \varkappa^2,\ p_2^2 = \varkappa^2,\ (p_1 - p_2)^2 = \mu^2) = gZ^{-1}\gamma_5 = g_1\gamma_5\,, \quad (i\gamma p_1 + \varkappa = i\gamma p_2 + \varkappa = 0),$$

for the renormalization of the coupling constant in meson theory. The « bound state » contribution then involves the (unknown) renormalized coupling constant g_1, and has the form,

$$\text{Im } M_{R\alpha\beta}(\lambda'\lambda) = -\left(\frac{g_1^2}{4\pi}\right)\cdot 4\pi^2\tau_\alpha\tau_\beta\{u^{\lambda'}(i\gamma k)u^{\lambda}\}\delta(\mu^2 + 2P^2 - 2k_0p_0)\,.$$

The renormalized coupling constant defined from $\Gamma_5(p_1 = p_2) = g_1'\gamma_5$ then differs from the above by $O(\mu^2/\varkappa^2)$.

After the work described here was completed, the authors, learnt from Prof. M. L. GOLDBERGER that he has obtained identical results in collaboration with Professors NAMBU and OEHME, and is applying the relations to calculate S phase shifts.

RIASSUNTO (*)

Si generalizzano le considerazioni di un precedente lavoro per scrivere le relazioni dello scattering non in avanti, e della dispersione per inversione di spin. Si discute il contributo dello « stato legato ».

(*) *Traduzione a cura della Redazione.*

Reprinted from THE PHYSICAL REVIEW, Vol. 112, No. 1, 283-287, October 1, 1958

Relativistic Field Theory of Unstable Particles

P. T. MATTHEWS AND ABDUS SALAM
Imperial College, London, England
(Received June 9, 1958)

An attempt is made to clarify the fundamental physical basis of the intrinsic characteristics of unstable elementary particles and to replace the conventional, purely phenomenological, description in terms of decaying states and complex energy levels, by definitions which are consistent with general requirements of relativistic quantum mechanics like Hermiticity, unitarity, and causality.

1. INTRODUCTION

OF the sixteen known "elementary" particles, all but four are unstable. In spite of this, no theory of unstable particles exists which is consistent with the general properties of relativistic quantum mechanics. This is particularly unfortunate since most of the recent significant progress in elementary particle theory has been made through the application of these general principles.

The conventional treatment of decays is a kind of "adiabatic" approximation. The particle is first tacitly assumed to have a definite mass, M. The rate of decay per unit time, λ, is calculated like any transition between stable particles. A beam of such particles will then be attenuated according to the relation

$$dn(\tau)/d\tau=-\lambda(M)n(\tau), \tag{1.1}$$

where $n(\tau)$ is the number of particles in the beam at proper time τ. This is interpreted as the individual particles of the beam having a wave function in their own rest system,

$$\psi(\tau)=\exp[+iM\tau-\tfrac{1}{2}\lambda|\tau|]. \tag{1.2}$$

Although this scheme is generally accepted as giving the "right answer" if the interaction producing the decay is sufficiently weak, there exists no exact theory to which this is supposed to be an approximation. Furthermore, none of the assumptions made are consistent with a complete theory of elementary particles. It is intuitively clear that an unstable particle does not have a definite mass, since it does not live long enough for an exact measurement to be made. The wave function (1.2) is certainly a useful approximation, but the complex mass and consequent lack of conservation of probability of the wave function violate general features of any complete theory, and are typical of an approximate one-particle treatment of what is essentially a many-particle problem.

Our object is to propose a theory of unstable particles, which is consistent with general requirements of a relativistic quantum field theory. The aspect of an unstable particle on which we shall concentrate is the uncertainty in the mass. In a rough way, this is already inherent in the conventional treatment. The wave function (1.2) can be written in terms of its Fourier transform

$$\psi(\tau)=\frac{\lambda}{2\pi}\int\frac{\exp[-im\tau]}{(m-M)^2+\frac{1}{4}\lambda^2}dm, \tag{1.3}$$

which may be interpreted as describing a distribution of mass values, with a spread, Δm, related to the mean life $\Delta\tau(=1/\lambda)$, by an uncertainty relation

$$\Delta m\Delta\tau\sim 1, \tag{1.4}$$

or

$$\Delta Q\Delta\tau\sim 1,$$

where ΔQ is the uncertainty in the Q value of the decay. With this as starting point, we build in subsequent sections a local field theory of unstable particles which can adequately describe their "static" characteristics, like mean mass and mean life. In the course of this development we shall give a very direct physical interpretation to the spectral function introduced by Källén[1] and Lehmann.[2,3]

Nishijima and Zimmermann[4] have recently shown that a local field theory can be set up for a stable bound state which is similar to that for stable "elementary" particles so far as the axiomatic formulation is concerned. By the same token the theory of unstable particles we propose here incorporates the theory of unstable bound states. These appear as resonances in scattering processes.

[1] G. Källén, Helv. Phys. Acta 25, 417 (1957).
[2] H. Lehmann, Nuovo cimento 2, 347 (1954). See, also, M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954).
[3] We remark that the only previous definition of lifetime within the general framework of conventional theory is that of R. E. Peierls, *Proceedings of Glasgow Conference on Nuclear and Meson Physics* (Pergamon Press, Inc., London, 1954), p. 296 and B. Zumino, New York University Research Report *CX-23*, 1956 (unpublished). Both these authors surmise that the mean mass and mean life of unstable particles are determined by the real and imaginary parts of complex poles of the propagator. Since this function contains branch points, the "physical" sheet is defined by a cut along the real axis. It has been shown by Bogoliubov *et al.* [Bogoliubov, Mednedev, and Polivanov, "Problems of the theory of dispersion relations," mimeographed lecture notes, Institute for Advanced Study, Princeton, 1957 (unpublished)] that the single function $\Delta(p^2)$, which defines the various propagation functions, has no singularities anywhere in the physical sheet of the p^2 plane, except on the positive real axis. Peierls has surmised that poles which give the mean mass and lifetime of an unstable particle lie on the "unphysical" sheets.
[4] K. Nishijima, Phys. Rev. 111, 995 (1958); W. Zimmermann (to be published).

2. DEFINITIONS

Consider a particle ϕ which decays into two stable particles χ, each of mass m. We shall assume that these particles can be described within the context of relativistic quantum theory. That is to say, we assume[5]

(a) the possibility of describing them by linear field operators $\phi(x)$ and $\chi(x)$;

(b) the transformation laws of these fields under the transformations of the inhomogeneous Lorentz group;

(c) the causality condition, which asserts the commutativity of field operators with space-like separation;

(d) the asymptotic condition *for the stable particles only*.

For simplicity consider ϕ to be a Bose particle.

We assert that the properties of the unstable particle ϕ are determined by a density matrix

$$\rho=\sum|p_\mu,2\chi\rangle\frac{\rho(p^2)}{p^2}\langle p_\mu,2\chi|, \tag{2.1}$$

where $|p_\mu,2\chi\rangle$ are a (normalized) set of states of energy momentum p_μ, which asymptotically are the two χ states into which the ϕ can decay. Since the χ particles satisfy the asymptotic condition, these states can be formed in the conventional way. Ideally $\rho(p^2)$ determines the probability of a Q value corresponding to $(p^2)^{\frac{1}{2}}$ being observed in the decay. This point is discussed further in Sec. 4.

The mean mass of the ϕ particle is defined as the mean value of the (mass)^2 operator for the distribution

$$M^2=\frac{\mathrm{Tr}[P^2\rho]}{\mathrm{Tr}[\rho]}=\int\rho(p^2)d^4p\Big/\int\frac{\rho(p^2)}{p^2}d^4p$$

$$=\int p^2\rho(p^2)dp^2\Big/\int\rho(p^2)dp^2. \tag{2.2}$$

The last equality holds since $\rho(p^2)$ vanishes unless the vectors p_μ are time-like. By definition,

$$\rho(p^2)=0 \quad \text{for} \quad p^2<4m^2. \tag{2.3}$$

Thus

$$M^2>4m^2.$$

The inverse lifetime is determined by the spread of the mass values about the mean.

$$\left(\frac{M\lambda}{2\pi}\right)^2=\frac{\frac{1}{2}\,\mathrm{Tr}[\{(P^2)^2-(M^2)^2\}\rho]}{\mathrm{Tr}[\rho]}$$

$$=\frac{1}{2}\int(p^2-M^2)^2\rho(p^2)dp^2\Big/\int\rho(p^2)dp^2. \tag{2.4}$$

To relate ρ to the field operator ϕ, we make the basic assumption of our theory that

$$\theta(p_0)\rho(p^2)d^4p=(2\pi)^3\sum_{p,p+dp}|\langle 0|\phi(0)|p_\mu,2\chi\rangle|^2, \tag{2.5}$$

where

$$\theta(p_0)=1, \quad p_0>0$$
$$0, \quad p_0<0.$$

Note that the summation is only over 2χ states, but includes all the variables required to define these states, in addition to the energy and momentum.

For an unstable fermion of spin one-half, the final states must carry an additional label, which denotes the spin direction and particle-antiparticle character of the final state. We denote this variable explicitly by α, β. Then the unstable particle is described by a density matrix

$$\rho=\sum|p_\mu,\alpha\rangle\frac{\rho_{\alpha\beta}(p_\mu)}{p^2}\langle p_\mu,\beta|. \tag{2.6}$$

We can write the spinor matrix

$$\rho_{\alpha\beta}(p_\mu)=[\rho(p)]_{\alpha\beta}, \tag{2.7}$$

where

$$p=i\gamma\cdot p. \tag{2.8}$$

The mass operator[6] is now $i\gamma\cdot P$ and the mean mass is thus determined by

$$M=\frac{\mathrm{Tr}[i\gamma\cdot P\rho]}{\mathrm{Tr}[\rho]}$$

$$=\int\mathrm{tr}[p\rho(p)]dp^2\Big/\int\mathrm{tr}[\rho(p)]dp^2, \tag{2.9}$$

where tr[ ] denotes the spinor trace. The inverse lifetime is again determined by the spread in mass values,

$$\left(\frac{\lambda}{2\pi}\right)^2=\frac{\frac{1}{2}\,\mathrm{Tr}[(P^2-M^2)\rho]}{\mathrm{Tr}[\rho]}$$

$$=\frac{1}{2}\int\mathrm{tr}[(p-M)^2\rho]dp^2\Big/\int\mathrm{tr}[\rho]dp^2. \tag{2.10}$$

If the particle is described by a local, causal, relativistic field $\psi(x)$, we postulate that

$$\theta(p_0)\rho_{\alpha\beta}(p)d^4p$$

$$=-(2\pi)^3\sum_{p,p+dp}\frac{\langle 0|\psi_\alpha(0)|p_\mu\chi_1\chi_2\rangle\langle p_\mu\chi_1\chi_2|\psi_\beta^*(0)|0\rangle}{p_0}.$$

That this expression has the required transformation properties can be seen by comparing it with the expression for the number density.[7]

[5] Lehmann, Symanzik, and Zimmermann, Nuovo cimento 1, 205 (1955). A. S. Wightman, preprint of Lille Conference talk, 1957 (unpublished).

[6] Abdus Salam, Nuclear Phys. 5, 687 (1958).

[7] See, for example, A. S. Wightman and S. S. Schweber, Phys. Rev. 98, 812 (1955). The Bose density may be cast in a similar form,

$$\rho(p^2)=(2\pi)^3\sum_{p,p+dp}\langle 0|\phi(0)|p,2\chi\rangle\langle p,2\chi|i\dot{\phi}(0)|0\rangle/p_0.$$

3. RELATION TO CONVENTIONAL THEORY

To make the discussion explicit, we shall consider in this section a model particle ϕ whose only interaction is that which causes its decay into two χ particles. The χ particles have no other function except to produce the ϕ particle when they come together. The ϕ, once created, can only decay.

For simplicity let us suppose that a selection rule exists so that the field ϕ only has nonvanishing matrix elements between the vacuum and 2χ states. Then the ρ defined by (2.5) can be identified with the ρ of Lehmann.[2] If we try to set the theory up in Lagrangian form, we may expect from such a Lagrangian an equation of motion of the form

$$\Box\phi=\alpha\phi+F[\chi], \tag{3.1}$$

where F is some functional of the field χ referring to the stable χ particle of mass m. Then a theorem of Lehmann[2] states that for such an equation, irrespective of the actual form of F,

$$\alpha=\int p^2\rho(p^2)dp^2\Big/\int\rho(p^2)dp^2. \tag{3.2}$$

This was just our definition of M^2, and so we have

$$(\Box-M^2)\phi=F[\chi]. \tag{3.3}$$

For $F[\chi]$ we make the usual assumption

$$F=g\chi^2. \tag{3.4}$$

In the limit $g\to 0$, we have

$$\rho=\delta(p^2-M^2). \tag{3.5}$$

Lehmann's theorem[2] states that the effect of the decay interaction for the model particle is to spread this δ function into a finite mass distribution about the same mean value. To obtain an approximate expression for the small g, we make use of the exact expression

$$\rho(p^2)=(1/2\pi)\Delta^{1\prime}(p^2)=(1/\pi)\,\mathrm{Im}\Delta_c'(p^2), \tag{3.6}$$

where

$$\Delta_c(p^2)=(p^2-M^2+i\epsilon)^{-1}. \tag{3.7}$$

According to Dyson,[8] $\rho(p^2)$ has the form of a "Cauchy" distribution with

$$\rho(p^2)=\frac{I(p^2)}{[p^2-M^2-R(p^2)]^2+I^2(p^2)}, \tag{3.8}$$

where

$$I(p^2)=g^2\theta(p^2-4m^2)(1+4m^2/p^2)^{\frac{1}{2}}/\pi, \tag{3.9}$$

if one approximates to order g^2. This expression has a mean value approximately at M^2, provided $M^2\gg 4m^2$ and g^2 is small. The second moment of a Cauchy distribution is mathematically infinite, but it is clear that with any reasonable definition the spread of the distribution is determined by $I(p^2)$. It has been suggested by Källén[9] that the correct form for $\rho(p^2)$ may be a series of the nature of an exponential, in which case (3.8) would be replaced by

$$\rho(p^2)\sim\frac{1}{I(p^2)}\exp\left[-\left(\frac{p^2-M^2}{I(p^2)}\right)^2\right]. \tag{3.10}$$

With this form the second moment would exist[10] and give a lifetime, according to (2.4), in agreement with the conventional calculation.

The fact that perturbation theory takes the zeroth approximation to ρ to be a δ function shows that the limit $g\to 0$ is certainly not a straightforward mathematical procedure, and it is clear that if the correct form for (3.8) is in the nature of (3.10) an expansion in powers of g^2 is not valid under any circumstances.

4. PHYSICAL INTERPRETATION OF ρ

To give a complete theory of unstable particles it would be necessary to consider not only the "static" properties such as mean mass and mean life, but also the dynamics of the particle interactions—its production and scattering. In principle, these two aspects cannot be separated since conceptually the end products of any complete interaction must be regarded as stable. (One can only discuss the production of unstable particles in the approximation in which they are regarded as stable.) To treat unstable particles on an exact basis, it is necessary to consider both the production and the decay process together. The density ρ through which we have defined the mean mass and mean life is related to the probability distribution of the Q values of the decay products in a production and decay chain, such as $\pi^-+p\to\kappa^0+\Sigma^0\to\kappa^0+\Lambda^0+\gamma$. (For this discussion we treat K as stable.) The experimenter observes the Q values of the $\Lambda+\gamma$. These show a sharp peak (due to "real" Σ's) and a diffuse spread (due to "virtual" Σ's, or direct Λ production with radiative γ's from the original charged particles). We shall assume that these parts of the Q spectrum can be clearly separated. Our ρ is closely related to the shape of the sharp peak (which has a width $\sim$10 kev due to the uncertainty principle, corresponding to the mean life $\sim 10^{-19}$ sec). In order to give a precise definition of ρ in terms of this shape, it is necessary to extract the dependence of the shape on the particular production process, which has been selected. This, together with the dynamic properties

[8] We have used the relations $\Delta_c'(p^2)=[p^2-M^2-\Sigma^*(p^2)]^{-1}$, and define $R(p^2)+iI(p^2)=\Sigma^*(p^2)/\pi$.

[9] G. Källén, *Proceedings of the* CERN *Symposium on High-Energy Accelerators and Pion Physics, Geneva, 1956* (European Organization of Nuclear Research, Geneva, 1956), Vol. 2, p. 187.

[10] If one introduces a variable s, canonical to the mass variable p^2, and defines $G(s)=\int\exp(ip^2s)\rho(p^2)ds$, then $\mathrm{Tr}[\rho]=G(0)$, $\mathrm{Tr}[P^2\rho]=-iG'(0)$, etc. Since $\rho(p^2)=0$ for $p^2<4m^2$, $G(s)$ is the boundary value of an analytic function in the upper-half $s(=s_1+is_2)$ plane. The existence of higher moments of $\rho(p^2)$ is connected with the analyticity of $G(s)$ at $s=0$. Note that the variable s is not the same as the explicit proper time variable $\tau=x^2$ which appears in the Fourier transform, $\Delta(\tau)=(2\pi)^{-4}\times\int\Delta(k)e^{ik\cdot x}d^4k$.

FIG. 1. The Feynman diagram of pure resonance scattering of two χ particles through the ϕ-particle resonance.

of the unstable-particle problem, will be considered in a separate paper.

If the lifetime of the particle (or bound state) is even shorter than the Σ^0 lifetime, the distinction between the sharp and diffuse parts of the Q spectrum become obscured. In this case the particle shows itself most clearly as a resonance. In the terminology of the previous section, we may consider the scattering of the decay products—the two χ particles. The total cross section is related by the optical theorem to the imaginary part of the forward scattering amplitude. If this may be assumed to be purely resonance scattering with the ϕ particle as intermediate state, the scattering amplitude is proportional to the ϕ-particle propagator (Fig. 1). The imaginary part is just ρ, apart from numerical factors. Thus within the severe limitations of the approximation, the total cross section provides a measure of ρ. Even in the case of a relatively long-lived particle the consideration of the resonance in the elastic scattering of the decay products may provide the clearest theoretical means to isolate ρ. Thus, conceptually, ρ for neutron decay would be given most clearly by the spread of the so-called "bound-state" term in the conventional $\pi+p$ scattering dispersion relations.

5. DISCUSSION

A more detailed analysis of the structure of unstable particles would require information on the higher moments of the distribution. As long as the particle is described by its first two moments only—the mean mass and mean life—the conventional treatment in terms of a complex mass provides an adequate phenomenological description.

To bring the discussion of the model particle nearer to reality, let us suppose that it has two interactions: a weak decay interaction into two particles of mass m_w, specified by a coupling constant g_w; and a strong interaction with particles of mass m_s, $(M^2<4m_s^2)$, and coupling g_s. The complete spectral function is $\rho(g_s,g_w)$. The forms of ρ in the various approximations of letting one or other, or both, of these interactions become zero are shown in Fig. 2. M_0 is the so-called "bare" mass and the forms of ρ follow directly from the discussion of Lehmann[2] and its extension given below.[11]

[11] It may be necessary to modify the definitions of mass and life time to conform more closely with experimental practice. One simple procedure is to limit the integrations in Eqs. (2.2) and (2.4) and set the upper limit $=4m_s^2$. A discussion for the realistic cases will be published elsewhere.

It will be remarked that from the point of view proposed here, stable and unstable particles are treated on a precisely equal footing. The basic quantities are the fields and their associated spectral functions. The notion of a particle is a qualitative one which is related to the peaks in the spectral function. If the width of the peak is of the order of the π-meson mass, the mean life is approximately 10^{-23} sec, but the mass spread of the "particle" may then be observed through the resonance in the scattering of the particles into which it decays. (The "33" resonance of the π-nucleon system is a

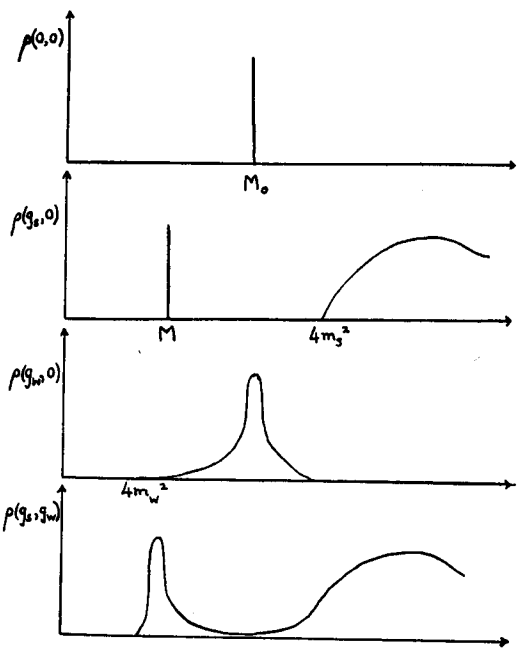


FIG. 2. A schematic drawing of the forms of the spectral function of a strongly produced (g_s) and weakly decaying (g_w) particle in various limits in which the different couplings are taken to be zero. The heavy lines denote δ functions. Note that any sharp peak in ρ may usefully be interpreted as an unstable particle, and in this way a single field may represent more than one unstable particle.

"particle" of this type.) For mean lives greater than 10^{-15} sec (corresponding to mass uncertainties of less than 1 ev) the mean life is experimentally the more accessible quantity, and the fundamental uncertainty in the mass is masked by experimental errors. The stable particles are those peaks in the spectral functions which can be approximated by δ functions. Postulating the appearance of a δ function in ρ, corresponding to each stable particle, is equivalent to the existence of asymptotic fields in the axiomatic approach.

From the present point of view the existence of several particles may then be subsumed in a single

field, and the problem of elementary particle physics is shifted from the question of the number of elementary particles to the number of elementary fields. It would appear that the future task of fundamental theory would be to look for criteria which specify elementary fields.

ACKNOWLEDGMENTS

The authors would like to thank many people for stimulating discussions and some healthy scepticism, particularly Professor J. R. Oppenheimer, Professor R. E. Peierls, and members of the high-energy physics groups in Edinburgh and Imperial College, London.

Reprinted from THE PHYSICAL REVIEW, Vol. 115, No. 4, 1079–1084, August 15, 1959

Relativistic Theory of Unstable Particles. II

P. T. MATTHEWS AND ABDUS SALAM
Imperial College, London, England
(Received March 16, 1959)

This paper is a direct continuation of an earlier paper (I) where an attempt was made to set up a field-theoretic foundation for the theory of mean mass and lifetime of an unstable particle. It was argued in I that the decay-time plot of a beam of unstable particles is a concept peculiar to a single-particle theory; that from a field-theoretic point of view, mass (the variable conjugate to proper time) rather than time has the primary significance. Here we show that the spectral function $\rho(m^2)$ appearing in the (field-theoretic) one-particle propagator has a direct significance as the probability of finding in production an unstable particle of mass m. This allows us to define a "one-particle" state for the unstable particle as a superposition of its outgoing decay states suitably weighted in mass space [with a factor which is the square-root of $\rho(m^2)$]. The proper-time propagation of this state gives the decay amplitude, and its modulus is ideally the experimentally observed decay-time plot.

The time plot is explicitly evaluated for π decay. Insofar as the distribution of mass values for the π meson starts with the μ mass (assumed stable), the time plot is not merely the conventional decay exponential $e^{-\tau/\tau_0}$. There are additional terms which become important about a hundred lifetimes after the particle is created.

Finally we compare the time plots for particle and antiparticle decays on the basis of CTP invariance.

1. INTRODUCTION

IN a previous paper[1] (referred to as I) some of the properties of an unstable particle have been formulated in a manner which is consistent with the general requirements of relativistic quantum mechanics. This paper is a direct continuation of I.

The main result of I was that one may associate local relativistic field operators with unstable elementary particles just as well as with stable particles; these fields satisfy the causality condition and possibly also satisfy local equations of motion. These latter may contain parameters corresponding to (bare) mass and charge. The only difference between the stable and the unstable case lies in the fact that fields corresponding to stable particles possess asymptotic limits, so that "in" and "out" fields can be defined.[2] This is not true for the unstable case. This statement is equivalent to the following:

(1) The set of "in" states (or "out" states) exist for stable particles only. States corresponding to unstable "particles" can be defined (see below) but these are not eigenstates of the mass-operator, P_μ^2, corresponding to a unique mass value for the unstable particle. Nor do such states appear directly in the completeness relations

$$\sum|\text{in}\rangle\langle\text{in}|=1, \tag{1.1}$$

or

$$\sum|\text{out}\rangle\langle\text{out}|=1. \tag{1.2}$$

(2) For the stable case the spectral function $\rho(\kappa^2)$ which appears in the definition of the Green's function

$$\Delta_c'(x-y)=i\langle T(\phi(x),\phi(y))\rangle_0$$

$$=(2\pi)^{-4}\int_0^\infty d\kappa^2\int\frac{\rho(\kappa^2)\exp[ip(x-y)]}{p^2-\kappa^2+i\epsilon}d^4p, \tag{1.3}$$

has the form

$$\rho(\kappa^2)=\delta(\kappa^2-M^2)+\sigma(\kappa^2). \tag{1.4}$$

The position of the δ-function singularity defines the physical mass, M, of the stable particle associated with ϕ. Equivalently $\Delta_c'(p^2)$ has a real pole at M^2. There is no such pole for the unstable case, nor does ρ contain a δ-function singularity of the above type.

The chief problem of the phenomenological field theory of unstable particles, then, is that of providing a suitable theoretical entity to be correlated with the experimentally observed quantities like mean mass, mean lifetime, partial lifetimes, etc. This was attempted in I, where it was suggested that the spectral function ρ, defined above, provides a suitable basis for such a theoretical development. Here we start by substantiating the claim made in I that $\rho(\kappa^2)$ gives the probability of producing an unstable particle of mass κ^2. This allows us to define for such particles a one-particle state, $|s\rangle$, as a linear superposition in mass of "out" states of its stable decay products. In this superposition all "out" states of mass κ are weighted with a factor which is essentially the square root of $\rho(\kappa^2)$. Clearly the state $|s\rangle$ is not an eigenstate of the mass-operator p_μ^2. However its proper-time development gives the probability that the particle has not decayed up to a certain time, thus yielding the results of an attenuation experiment with a beam of unstable particles. This is the quantity of primary experimental interest for all long-lived particles (mean lives $>10^{-16}$ sec).

Finally we prove the exact equality of particle and antiparticle lifetimes on the basis of the present formulation.

2. THE SPECTRAL FUNCTION

Before considering unstable particles specifically we list some properties of spectral functions, ρ, for a real Bose field ϕ. Let $|i\rangle_{\text{in}}$ and $|i\rangle_{\text{out}}$ be a set of states congruent to the state $\phi|0\rangle$. By "congruent" we mean those states which are coupled to the state $\phi|0\rangle$ through

[1] P. T. Matthews and Abdus Salam, Phys. Rev. 112, 283 (1958).
[2] Lehmann, Symanzik, and Zimmermann, Nuovo cimento 1, 205 (1955).

the interaction. The label $i=1, 2, \cdots$ gives the distinct type of such states; for example states congruent to $\pi^+|0\rangle$ are states $|e^+\nu\rangle$, $|\mu^+\nu\rangle|\pi^+\pi^0\pi^0\rangle$, $|p\bar{n}\rangle$, $\cdots$ etc. of arbitary energy and momentum. Define

$$\rho_i(p^2)=\sum|\langle 0|\phi(0)|i,p\rangle|^2. \tag{2.1}$$

The only states appearing in the sum are i-states of total energy-momentum p. To make the definition more precise, consider the case when i denotes a particular two-particle "out" state,

$$|i\rangle_{\text{out}}=|k_1,k_2\rangle_{\text{out}}, \tag{2.2}$$

where k_1, k_2 are the four-momenta of the two particles. The state $|k_1,k_2\rangle_{\text{out}}$ is defined by the prescriptions of Lehmann, Symanzik, and Zimmermann.[2] Using the covariant normalization

$$\langle k_1|k_2\rangle\Delta^+(k_1)=(2\pi)^3\delta(k_1-k_2), \tag{2.3}$$

where

$$\Delta^+(k)=\theta(k_0)\delta(k^2-m^2), \tag{2.4}$$

Eq. (2.1) reads

$$\rho_i(p^2)=(2\pi)^{-3}\int|\langle 0|\phi(0)|k_1,k_2\rangle|^2\Delta^+(k_1)\times\Delta^+(k_2)\delta(p-k_1-k_2)d^4k_1d^4k_2. \tag{2.5}$$

Define

$$\rho(p^2)=\sum_i\rho_i(p^2). \tag{2.6}$$

From the completeness relation (1.1), it is clear that

$$\rho(p^2)=\tfrac{1}{2}(2\pi)^{-3}\int\langle 0|\{\phi(x),\phi(0)\}|0\rangle\exp[ipx]d^4x$$
$$=(1/\pi)\,\text{Im}\Delta_c'(p^2). \tag{2.7}$$

The expression above determines ρ in terms of Δ_c'. Alternatively each ρ_i can be expressed in terms of Δ_c' and Γ_i, where Γ_i is the appropriate (proper) "vertex part" allowing for a transition from a ϕ-state to the state $|i\rangle$. Specializing again to a particular two-particle state,

$$\int\langle 0|\phi(x)|k_1,k_2\rangle\exp[ipx]d^4x$$
$$=(2\pi)^4\langle 0|\phi(0)|k_1,k_2\rangle\delta(p-k_1-k_2)$$
$$=(2\pi)^4\Delta_c'(p)\Gamma_i(p-k_1-k_2)\delta(p-k_1-k_2). \tag{2.8}$$

Here Γ_i is the "proper" vertex part defined by Dyson,[3] with external lines corresponding to the ϕ field and k_1, k_2 particles. In any field theory with a local interaction, the validity of this relation can be proved immediately by using Schwinger's functional differentiation techniques.[4] Then

$$\rho_i(p^2)=(2\pi)^{-3}\int|\Delta_c'(p)\Gamma_i(p,k_1,k_2)|^2\times\delta(p-k_1-k_2)\Delta^+(k_1)\Delta^+(k_2)d^4k_1d^4k_2. \tag{2.9}$$

If we define

$$I_i(p^2)=\pi\int|\Gamma_i(p,k_1,k_2)|^2\times\delta(p-k_1-k_2)\Delta^+(k_1)\Delta^+(k_2)d^4k_1d^4k_2, \tag{2.10}$$

then

$$\rho_i(p^2)=(1/\pi)|\Delta_c'(p^2)|^2I_i(p^2). \tag{2.11}$$

It is crucial for subsequent work that the δ-functions appearing in the integrand ensure that

$$I(p^2)=0 \quad \text{for} \quad p^2=(k_1+k_2)^2\leqslant 4m^2=M_i^2. \tag{2.12}$$

We shall refer to M_i as the threshold mass for the i state. Define

$$I=\sum_i I_i. \tag{2.13}$$

From (2.6), (2.7), and (2.11),

$$\text{Im}\Delta_c'/|\Delta_c'|^2=I(p^2). \tag{2.14}$$

Equation (2.14) shows quite generally that I is the imaginary part of $(\Delta_c')^{-1}$:

$$\Delta_c'=[X-iI]^{-1}, \tag{2.15}$$

where X is real. Using (2.7), ρ must have the form

$$\rho_i(p^2)=-\frac{1}{\pi}\frac{I_i(p^2)}{X^2(p^2)+I^2(p^2)}. \tag{2.16}$$

Rewrite

$$X(p^2)=p^2-M_0^2-R(p^2), \tag{2.17}$$

where M_0 is the bare mass of the ϕ field. $R(p^2)$ and $I(p^2)$ can now be recognized as the real and imaginary parts of the proper self-energy part. Then

$$\rho_i(p^2)=-\frac{1}{\pi}\frac{I_i(p^2)}{[p^2-M_0^2-R(p^2)]^2+[\sum_i I_i(p^2)]^2}. \tag{2.18}$$

(The similarity of this expression to the familiar single-level formulas in nuclear resonance theory is clear.) The result is remarkable insofar as all ρ_i have a common denominator. This is the main result of this section.

Before concluding this section note Lehmann's mean-mass theorem[5] for the bare mass,

$$\int(p^2-M_0^2)\rho(p^2)dp^2=0. \tag{2.19}$$

[3] F. J. Dyson, Phys. Rev. 76, 1736 (1949).

[4] J. Schwinger, Proc. Natl. Acad. Sci. U. S. 37, 452 (1951).

[5] H. Lehmann, Nuovo cimento 2, 347 (1954), where the precise conditions for the validity of this theorem are shown.

This applies to all particles (stable or unstable). A stable theory is one for which $X(p^2)$, (2.17), has a real zero M_s^2,

$$M_s^2-M_0^2-R(M_s^2)=0, \tag{2.20}$$

such that

$$M_s^2<M_{s1}^2.$$

Here M_{s1} is the threshold mass of the state of lowest mass congruent to the state $\phi|0\rangle$. So defined, M_s is the physical mass of the stable ϕ particle. For such a theory,

$$\rho(p^2)=\delta(p^2-M_s^2)+\sigma(p^2), \tag{2.21}$$

where

$$\sigma(p^2)=0 \quad \text{for} \quad p^2\leqslant M_{s1}^2.$$

This implies that an equivalent definition of the physical mass is given by

$$\int^{M_{s1}^2}(p^2-M_s^2)\rho(p^2)dp^2=0, \tag{2.22}$$

which is to be compared with (2.19) for the bare mass. This is an important definition, which is used below.

3. THE MASS PLOT

We now wish to specialize to the case of realistic unstable elementary particles. All these have the characteristic that the interactions through which they are produced are stronger than those through which they decay, and are referred to as "strong" and "weak" interactions, respectively. If the weak interactions could be switched off, the particle would be stable. As discussed in I, this implies that if only the strong interactions are considered, ρ_s, the approximate expression for ρ, would have the form (2.21), where M_s now denotes what we call the strong mass. The effect of the weak interactions is to spread the δ-function into a finite distribution, and to shift the mean somewhat from M_s^2.

We now show that for unstable elementary particles with weak and strong interactions, the spectral functions $\rho_i(\kappa^2)$, for all i such that $M_i<M_s$, give to a good approximation the probability of the particle being produced with a mass κ and decaying through the i mode. Here M_i is the threshold mass for the i decay mode and

$$\rho_i(\kappa^2)=0 \quad \text{for} \quad \kappa^2<M_i^2. \tag{3.1}$$

Suppose that the matrix element for the production of the unstable particle, four-momentum p, in a certain process, is $F(p^2)$. If the particle subsequently decays via the mode "i" into k_1 and k_2, using Schwinger's method, it is easy to show that the amplitude for the whole process is

$$F(p^2)\Delta_c'(p)\Gamma_i(p,k_1,k_2)\delta(p-k_1-k_2), \tag{3.2}$$

provided one can neglect the effect of other (virtual) particles which in graphical language link the factor F to the vertex Γ_i. The probability of observing decay products corresponding to a particular value of p^2 for the unstable particle is then proportional to

$$P(p^2)=\int|F(p^2)\Delta_c'(p)\Gamma_i(p,k_1,k_2)|^2$$

$$\times\delta(p-k_1-k_2)\Delta^+(k_1)\Delta^+(k_2)d^4k_1d^4k_2. \tag{3.3}$$

We call this the mass plot. If it is assumed that F varies slowly with (p^2) compared to the remaining factor, which is sharply peaked about the mean mass M ($\simeq M_s$) of the unstable particle (to be defined below more precisely), it can be factored out from this expression. The probability of mass p^2 for a given production process and i decay mode is then, by (2.11),

$$P(p^2)\simeq F(M^2)\rho_i(p^2), \tag{3.4}$$

which establishes the result. The approximation employed implies that we are neglecting the effects of the production mechanism on the properties of the unstable particle.

This is the stage to introduce a number of conventions to conform to experimental practice. Let us define weak thresholds to be those threshold masses M_j which satisfy

$$M_1, \cdots, M_n\leqslant M_s<M_{s1}, \tag{3.5}$$

where M_{s1} is now the threshold mass of the state of least mass which is congruent to the state via the strong interactions.[6] Define

$$\rho_{Dj}(p^2)=\rho_j(p^2)\theta(M_{s1}^2-p^2), \tag{3.6}$$

and

$$\rho_D(p^2)=\sum_{j=1}^{n}\rho_{Dj}(p^2) \tag{3.7}$$

We now define the mean mass of the unstable particle and the mass shift in terms of these expressions. From (2.19) and (2.21), which may be rewritten, in an obvious notation for the strong mass, as

$$\int(p^2-M_s^2)\rho_{sD}(p^2)dp^2=0, \tag{3.8}$$

it seems reasonable to adopt as the definition for the mean mass, M, of an unstable particle the relation

$$\int(p^2-M^2)\rho_D(p^2)dp^2=0. \tag{3.9}$$

This definition of mean mass, in terms of the truncated spectral function, ρ_D, conforms precisely to experimental procedure. As is clear from (2.18) and (2.12), this

[6] If there are any thresholds lying between M_s and M_{s1}, these should ideally be included as weak thresholds. It is experimental practice, however, not to associate these with the decay of a ϕ particle. A specific example is $\pi^+\to\mu^++\nu+\pi^0$. On the basis of the ideas presented here, one particle in 10^{12} of the so called π^+ structures would be produced with enough energy to decay in this fashion. However, to simplify the discussion we neglect all such cases.

isolates from the complete ρ function the peak in the neighborhood of M_s^2, (see Fig. 2 of I), with width given by the I_j, corresponding to what are conventionally regarded as the physically possible decay modes of the particle.

It is to be noticed also from (2.18) that in the limit of the weak coupling tending to zero ($I \to 0$), ρ_D becomes a δ-function with argument equal to the real part of the denominator. Thus, by (2.20), (3.8) is a restatement of the usual definition of the renormalized mass, when only strong interactions are considered.

The mass shift due to weak interactions is then defined to be $|M-M_s|$. The best example of such a mass shift is the case of K_1 and K_2 mesons. These have the same strong interactions and consequently the same strong mass. But the decay channels, and consequently the weak contributions to ρ_{K1} and ρ_{K2}, are different, leading to different lifetimes and a small difference in the mean mass values.

4. THE STATE VECTOR

Turning now to the problem of finding a suitable theoretical expression for the decay time plot, we first set up a state vector describing the unstable particle.

In I, the function ρ_D was used to define a density matrix, and the mean mass and lifetime were determined in terms of the mean value of the mass operator and its second moment for this density distribution. Since however ρ_D defines a density which is diagonal in the representation in which the operator P^2 is diagonal, it is possible to define a state (or more precisely states, one for each decay channel), which leads to the same expectation values of functions of P^2 as the corresponding mean values of ρ_D. The required state corresponding to channel j is[7]

$$|s_j\rangle = \int d\kappa \int |k_1,k_2\rangle_{\text{out}} \Delta^+(k_1)\Delta^+(k_2)_{\text{out}}\langle k_1,k_2|\phi(0)|0\rangle \times \delta((k_1+k_2)^2-\kappa^2)\theta(M_{s1}-\kappa)d^4k_1 d^4k_2. \quad (4.1)$$

It is easy to verify that

$$\langle s_j|s_j\rangle = \int \rho_{Dj}(\kappa^2)d\kappa^2. \quad (4.2)$$

The state $|s_j\rangle$ is thus an appropriately weighted linear superposition of the states into which the ϕ particle can decay. Define

$$|s_D\rangle = \sum_{j=1}^{n} |s_j\rangle, \quad (4.3)$$

then

$$\langle s_D|s_D\rangle = \int \rho_D(p^2)dp^2, \quad (4.4)$$

and

$$\langle s_D|P^2|s_D\rangle = \int p^2\rho_D(p^2)dp^2 \Big/ \int \rho_D(p^2)dp^2. \quad (4.5)$$

To get the (proper) time plot we must study the propagation of this state in (proper) time. The single-particle approximation has implicit in it the hypothesis that a single proper-time parameter can be attached to the decaying particle up to the instant of its decay. Writing[8]

$$i\frac{\partial|s_j(\tau)\rangle}{\partial\tau} = (P_s^2)^{\frac{1}{2}}|s_j(\tau)\rangle, \quad (4.6)$$

we get

$$|s_j(\tau)\rangle = \int d\kappa \exp[i\kappa\tau]\int |k_1,k_2\rangle_{\text{out}} \times\Delta^+(k_1)\Delta^+(k_2)_{\text{out}}\langle k_1,k_2|\phi(0)|0\rangle \times\delta((k_1+k_2)^2-\kappa^2)\theta(M_{s1}-\kappa)d^4k_1 d^4k_2. \quad (4.7)$$

The decay amplitude is

$$\mathcal{G}_j(\tau-\tau_0) = \langle s_j(\tau)|s_j(\tau_0)\rangle = \int \rho_{Dj}(\kappa^2)\exp[-i\kappa(\tau-\tau_0)]d\kappa^2, \quad (4.8)$$

and the partial decay probability for channel j is given by $|\mathcal{G}_j|^2$. It is satisfactory that since, by (2.18), the main κ dependence of ρ_{Dj} comes from the denominator which is the same for all j, essentially the same time plot is obtained if one studies a particular decay mode rather than all possible decay modes.

It is conventional to assume that $\mathcal{G}$ has exponential form. The relation of this assumption to the present theory is examined in the next section.

5. THE DECAY TIME PLOT

Consider, to be specific, the decay of a π meson into a μ meson and a neutrino. By (2.18) and (4.8), we have to evaluate

$$\mathcal{G}(\tau) = \int_{\mu^2}^{M_{s1}^2} \frac{I(\kappa^2)\exp[-i\kappa\tau]}{[\kappa^2-M_0^2-R_s(\kappa^2)-R_w(\kappa^2)]^2+I^2(\kappa^2)}d\kappa^2, \quad (5.1)$$

where μ is the μ-meson mass, R_s and R_w denote contributions arising from strong and weak interactions, respectively, and I is the imaginary part of the π-meson self-energy due to virtual decay into a μ meson. The square bracket can be closely approximated by

$$\kappa^2-m^2,$$

where m^2 is the strong mass[9] (including electromagnetic self-energy effects). Since the integrand is sharply

[7] States similar to these have been considered previously by R. F. Streater (private communication).

[8] This is the "single-particle" equation for a particle in its own rest frame.

[9] The mass m^2 corresponds to what is denoted by M_s^2 in Sec. 3.

peaked about this value, we may approximate by the expression[10]

$$\mathcal{G}(\tau)\simeq\int_{\mu^2}^{\infty}\frac{I(m^2)\exp[-i\kappa\tau]}{[\kappa^2-m^2]^2+I^2(m^2)}d\kappa^2. \tag{5.2}$$

The integral can be evaluated on the assumption that

$$I\ll m^2. \tag{5.3}$$

If the contour is closed in the first quadrant the enclosed pole gives rise to the usual exponential term. The integral from μ to $\mu+i\infty$, which has to be subtracted, is a Laplace transform. The result for large τ is

$$\mathcal{G}(\tau)\simeq\exp[-I\tau/2m]+\frac{2i\mu I}{\pi(m^2-\mu^2)^2}\frac{\exp[i(\mu-m)\tau]}{\tau}. \tag{5.4}$$

According to the conventional method of calculating lifetimes, the mean life τ_0 is given by

$$I/m=1/\tau_0. \tag{5.5}$$

The probability that the particle has not decayed after time τ is $|\mathcal{G}(\tau)|^2$ which besides the usual exponential contains additional corrections.

The appearance of correction terms is not a special feature of the approximations we have made, but an example of a general theorem, that the Fourier transform of any function which vanishes for all values of the argument less than some finite value, behaves for large τ like some power of τ. Thus the appearance of such a term in the time graph is a very specific prediction of our theory.[11] It arises directly from the very general property that the mass spectrum of the unstable particle is certainly zero for values of κ^2 less than the squared rest mass of the lightest decay products.

It is reasonable to assume that the correction terms are given correctly in order of magnitude by the above expressions for any particle. For most observed unstable particles the correction terms are only significant after a period of about one hundred lifetimes. For Σ^0, π^0, and Be^8 it may be as short as twenty to thirty life times. For these cases there may be effects of comparable magnitude which have been neglected in the derivation of (5.4).

6. EQUALITY OF PARTICLE AND ANTIPARTICLE LIFETIMES

By the procedure developed by Lehmann, Symanzik, and Zimmermann,[2] two sets of states can be formed for the stable particles, the "in" states and the "out" states. It is one of the assumptions of the theory that each of these sets is complete. Thus

$$\begin{aligned}1&=\sum|n\rangle_{\text{in in}}\langle n|\\&=\sum|n\rangle_{\text{out out}}\langle n|.\end{aligned} \tag{4.1}$$

If $\alpha, \beta, \cdots$ are a complete commuting set of operators, which commute with the S matrix, they can be used as labels to identify all the states in both these summations. Thus

$$\begin{aligned}1&=\sum_{\alpha',\beta',\cdots}|\alpha',\beta',\cdots\rangle_{\text{in in}}\langle\alpha',\beta',\cdots|\\&=\sum_{\alpha',\beta',\cdots}|\alpha',\beta',\cdots\rangle_{\text{out out}}\langle\alpha',\beta',\cdots|.\end{aligned} \tag{4.2}$$

If we take the $\langle\alpha'|\cdots|\alpha'\rangle$ matrix element of these equations, we obtain two alternative forms of the unit matrix in the subspace of all states corresponding to the eigenvalue α' of α. We denote this by

$$\begin{aligned}p^{(\alpha)}&=\sum_{\beta',\cdots}^{(\alpha)}|\beta',\cdots\rangle_{\text{in in}}\langle\beta',\cdots|\\&=\sum_{\beta',\cdots}^{(\alpha)}|\beta',\cdots\rangle_{\text{out out}}\langle\beta',\cdots|,\end{aligned} \tag{4.3}$$

where the summation is over all those states with a given value of α. In particular, taking α to be the total energy-momentum of the state,

$$\alpha=p_\mu,$$

the "in" and "out" states corresponding to this value give two alternative forms of the unit matrix for all states of this energy and momentum. From this it follows that $\rho(p^2)$ could equivalently have been defined in terms of the "in" states, or, in an obvious notation,

$$\rho_{\text{in}}(p^2)=\rho_{\text{out}}(p^2). \tag{4.4}$$

Now consider a non-Hermitian field ϕ representing an unstable particle and corresponding antiparticle. If we write symbolically for the particle

$$\rho_{\text{in}}(p^2)=\textstyle\sum_n|\langle 0|\phi(0)|n\rangle_{\text{in}}|^2, \tag{4.5}$$

then for the antiparticle

$$\bar{\rho}_{\text{in}}(p^2)=\sum_{\bar{n}}|\langle 0|\phi^\dagger(0)|\bar{n}\rangle_{\text{in}}|^2, \tag{4.6}$$

where $\bar{n}$ denotes the same state as n, but with all particles replaced by antiparticles. If the theory is invariant under change conjugation, C, these two expressions are equal. Even if the theory is not invariant under C but only under CTP, then

$$\begin{aligned}\langle 0|\phi(0)|n\rangle_{\text{in}}&={}_{\text{out}}\langle\bar{n}|\phi(0)|0\rangle\\&=\langle 0|\phi^\dagger(0)|\bar{n}\rangle_{\text{out}}{}^*.\end{aligned} \tag{4.7}$$

[10] We remark that in this approximation the time plot corresponding to a particular decay mode is, according to (2.18),

$$\mathcal{G}_i(\tau)=\frac{I_i(m^2)}{I(m^2)}\mathcal{G}(\tau),$$

as in the conventional treatment.

[11] Such inverse powers of τ in the decay amplitude were first discovered by M. Lévy in connection with calculations on the Lee model. We are greatly indebted to Professor Lévy for informing us of his results prior to publication, and also for pointing out the existence of the general theorem. See also G. Höhler, Z. Phys. 152, 546 (1958).

Thus

$$\rho_{\text{in}}(p^2)=\bar{\rho}_{\text{out}}(p^2). \tag{4.8}$$

Hence, by (4.4),

$$\rho(p^2)=\bar{\rho}(p^2). \tag{4.9}$$

Thus the time plots, as well as the mean mass and lifetime of particle and antiparticle, are the same.

Let us now consider partial lifetimes. Suppose that a particle has two alternative decay modes to be denoted by a and b. Define

$$\rho_{\text{in}}(a)=|\langle 0|\phi(0)|a\rangle|^2.$$

By the argument given above, it then follows that

$$\rho_{\text{in}}(a)=\bar{\rho}_{\text{out}}(\bar{a}),$$

where $\bar{\rho}(\bar{a})$ refers to the decay of the antiparticle into the corresponding antiparticle states. However the states a and b are clearly coupled through the weak interactions, which we write symbolically as

$$|a\rangle \equiv |b\rangle \text{ via weak.}$$

It then follows that

$$\rho_{\text{in}}(a)\neq\rho_{\text{out}}(a),$$

so that

$$\rho_{\text{in}}(a)\neq\bar{\rho}_{\text{in}}(\bar{a});$$

partial lifetimes of particle and antiparticle are not necessarily equal. However, if

$$|a\rangle \not\equiv |b\rangle \text{ via strong,}$$

the unit matrix can again be split, if weak interactions are neglected, into a part containing $|a\rangle$ and another part containing $|b\rangle$ and to first order in the weak interaction

$$\rho_{\text{in}}(a)=\rho_{\text{out}}(a).$$

Hence

$$\rho_{\text{in}}(a)=\bar{\rho}_{\text{in}}(\bar{a}).$$

That is to say, the partial lifetimes of particle and antiparticle are equal to first order in the weak interactions, provided the final states of the alternative decay modes are not coupled by strong interactions. A good example of this is the 2π and 3π decay modes of K^+ and K^-. These states are not coupled by the strong interaction since they have opposite G-parity (extended charge conjugation). Thus the $2\pi/3\pi$ branching ratio should be the same for K^+ and K^- to this approximation. However the 2π and 3π states are coupled through the electromagnetic interaction, and this coupling produces an inequality between the two branching ratios.

Finally we consider the partial lifetimes for the decay of a self-conjugate particle into two modes which are the particle conjugates of each other (for example assuming CP invariance $K_2^0\rightarrow\pi^0+e^{\pm}+\mu^{\mp}$). This is a special case of the partial lifetimes for particle and antiparticle considered above. From CTP invariance, the partial lifetimes are only equal to the extent that the final states are not coupled by strong or electromagnetic interactions.[12]

ACKNOWLEDGMENTS

The authors are indebted to Professor M. Lévy for very interesting and stimulating discussions, particularly on the question of the decay-time plot.

Note added in proof.—It has been pointed out to us by Professor Schwinger that the form of the correction terms given above is highly idealized since further restrictions on the mass plot are imposed by the experimenter, through the limitations of his measuring device. See the report of the 1959 International Conference on High Energy Physics, Kiev.

[12] These results are identical with those obtained by G. Lüders and B. Zumino [Phys. Rev. 106, 385 (1957)], who, however, base their definition of lifetime on the assumption of complex poles on the "unphysical sheet" of the propagator. This requires an analytic continuation which has not yet been defined. If such poles exist, they would provide a basis for alternative and equivalent definitions to those presented here. We would like to thank Dr. Zumino for a private communication on this point.

Reprinted from THE PHYSICAL REVIEW, Vol. 121, No. 1, 302–303, January 1, 1961

Nonleptonic Decay Modes of the Hyperons

G. FELDMAN,* P. T. MATTHEWS, AND A. SALAM
Imperial College, London, England
(Received July 11, 1960)

An attempt is made to correlate the observed asymmetries in hyperon decays and the hyperon decay times using a pole approximation to the decay matrix elements. It is found that a natural correlation is obtained if a near-global relation between pion coupling constants is assumed.

THE relation between the strong pion-nucleon coupling constant, the weak axial vector constant, and the leptonic decay rate of the pion was first obtained by Goldberger and Treiman.[1] This has recently been shown to arise directly from the pole term in the dispersion relation for the appropriate matrix element of the derivative of the axial vector current.[2] The argument can be generalized immediately to the leptonic decay modes of the strange particles.[3] We attempt here an interpretation of the nonleptonic decay modes of the hyperons, on the assumption that these too are dominated by the pole terms. These pole terms arise by considering the matrix elements as functions of the three external masses.

Our procedure is as follows. We assume the $\Delta I=\frac{1}{2}$ rule. We then show that the observed vanishing of the Σ asymmetry parameters α_+ and α_- (see below) strongly suggests global (or near global) symmetry of the pion couplings. With these strong-coupling constants and by choosing essentially one further parameter, we are able to correlate Λ,Σ decay times and asymmetries. Other authors[4] have considered the same problem by making use of specific global Lagrangians. The difference in our work lies in its stress on a "dispersion" rather than a Lagrangian approach and, in particular, in the inclusion of a K-meson pole pictured in Fig. 1(c).

The poles are represented graphically in Fig. 1. The strong-interaction vertex is by definition the corresponding renormalized coupling constant. We assume that the weak vertices satisfy $\Delta I=\frac{1}{2}$. The Y-N vertex is then of form

$$f(a_Y+b_Y\gamma_5), \quad (Y=\Sigma,\Lambda). \tag{1}$$

The K-π vertex is

$$f_K.$$

Introduce[5]

$$h_Y=fa_Y/(Y-N), \tag{2}$$

$$\sigma^2=\frac{a_\Sigma\,\Sigma+N}{b_\Sigma\,\Sigma-N}, \quad \lambda^2=\frac{a_\Lambda\,\Lambda+N}{b_\Lambda\,\Lambda-N}. \tag{3}$$

The matrix element[6] for $\Sigma^+\to n+\pi^+$ is

$$T_+=A^++\gamma_5B^+, \tag{4}$$

where

$$A^+=-(2g+g_{\pi\Sigma})\frac{h_\Sigma}{\sigma^2}-g_{\pi\Lambda}\frac{h_\Lambda}{\lambda^2}, \tag{5}$$

$$B^+=(2g-g_{\pi\Sigma})h_\Sigma-g_{\pi\Lambda}h_\Lambda. \tag{6}$$

Similarly for

$$\Sigma^-\to n+\pi^-,$$

$$A^-=\frac{g_{\pi\Sigma}h_\Sigma}{\sigma^2}-\frac{g_{\pi\Lambda}h_\Lambda}{\lambda^2}, \tag{7}$$

$$B^-=g_{\Sigma\pi}h_\Sigma-g_{\Lambda\pi}h_\Lambda-2F_\Sigma, \tag{8}$$

where

$$F_Y=f_Kg_{KY}/(K^2-\pi^2). \tag{9}$$

The matrix element T_0 for

$$\Sigma^+\to p+\pi^0$$

is determined by the $\Delta I=\frac{1}{2}$ rule to be

$$T_0=(1/\sqrt{2})(T_+-T_-). \tag{10}$$

If for the moment we neglect Σ,Λ differences (i.e., $\sigma=\lambda$, $h_\Sigma=h_\Lambda$), and assume $g_{\pi\Sigma}=g_{\pi\Lambda}$, then A^- vanishes identically.

If further

$$g=g_{\pi\Sigma}=g_{\pi\Lambda}, \tag{11}$$

then B^+ is also zero, and the corresponding asymmetry parameters α_+ and α_- vanish, in agreement with experiment. Since exact global symmetry is perhaps not realized in nature,[7] we use as exact experimental results $\alpha_+=\alpha_-=0$ to determine the renormalized pion constants. Thus

$$A^-=B^+=0 \tag{12}$$

* On leave from the Johns Hopkins University, Baltimore, Maryland.

[1] M. L. Goldberger and S. B. Treiman, Phys. Rev. 110, 1478 (1958).

[2] J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring (to be published). See also Y. Nambu, Phys. Rev. Letters 4, 380 (1960).

[3] J. Prentki (private communication).

[4] B. d'Espagnat and J. Prentki, Phys. Rev. 114, 1366 (1959); R. F. Sawyer, Phys. Rev. 112, 2135 (1958); S. A. Bludman, Phys. Rev. 115, 468 (1959); S. B. Treiman, Nuovo cimento 15, 916 (1960).

[5] Σ, Λ, N, K, π denote the masses of the corresponding particles and $g=g_{N\pi}$.

[6] We are assuming the $\Sigma-\Lambda$ parity even and $\Lambda-K$ parity odd.

[7] See A. Salam, Ninth Annual International Conference on High-Energy Physics, Kiev, July, 1959 (unpublished).

TABLE I. Deviations from global symmetry of pion coupling constants.

a_Σ/b_Σ	a_Λ/b_Λ	$g_{\pi\Sigma}/g$	$g_{\pi\Lambda}/g$
$(\Sigma+N)/(\Sigma-N)$	$(\Lambda+N)/(\Lambda-N)$	0.70	0.95
1	1	0.85	0.82
$(\Sigma-N)/(\Sigma+N)$	$(\Lambda-N)/(\Lambda+N)$	1	1

implies

$$g_{\Sigma\pi}=g\frac{2\sigma^2}{\sigma^2+\lambda^2},\quad g_{\Lambda\pi}=g\frac{h_\Sigma}{h_\Lambda}\frac{2\lambda^2}{\sigma^2+\lambda^2}. \tag{13}$$

Although these constants depend on the undetermined parameters a and b [Eq. (1)], nevertheless, they are insensitive to a reasonable range of values of a and b. See Table I.

Using the observed relation of decay rates,[8]

$$w_+=w_-, \tag{14}$$

one can now determine F_Σ, which in contrast to (13), is sensitive to the ratio a/b. Using this and the coupling constants (13), the amplitude for

$$\Lambda\to p+\pi^-$$

is

$$T_\Lambda=A_\Lambda+\gamma_5 B_\Lambda, \tag{15}$$

where

$$A_\Lambda=-\sqrt{2}\frac{g}{h_\Lambda}\left[\frac{h_\Lambda^2}{\lambda^2}+2h_\Sigma^2\frac{\lambda^2}{\sigma^2(\sigma^2+\lambda^2)}\right], \tag{16}$$

$$B_\Lambda=\sqrt{2}\left\{\frac{g}{h_\Lambda}\left[h_\Lambda^2-2h_\Sigma^2\frac{\lambda^2}{\sigma^2+\lambda^2}\right]+F_\Lambda\right\}. \tag{17}$$

Again some properties of Λ decay turn out to be insensitive to the choice of a and b given in Table I. The s-wave part A_Λ always lies within a few percent of A^0 (the s-wave part of $\Sigma^+\to p+\pi^0$) and the asymmetry α_Λ is always approximately equal to unity (within 5%) provided we choose F_Λ (i.e., $g_{K\Lambda}/g_{K\Sigma}$) to fit the observed Λ/Σ decay rates. It is remarkable that there exists a solution for F_Λ such that $(g_{K\Lambda}/g_{K\Sigma})^2$ lies consistently between 1 and 3 for all a and b. The value 3 is in agreement with the "pole" approximation to the $K^+ + N$ elastic scattering data.[9] Also, we find that $\mathrm{sign}(\alpha_\Lambda/\alpha_0) = -\mathrm{sign}(g_{K\Lambda}/g_{K\Sigma})$. See also d'Espagnat and Prentki.[4]

We can state our conclusions as follows. We use the "pole approximation" to the nonleptonic decays and assume a $\Delta I=\frac{1}{2}$ rule in the weak interaction. Then as a consequence of $\alpha_+=\alpha_-=0$ we find an almost global relation [Eq. (13)] amongst the coupling constants. The result is true for all choices of a and b. Next we use $w_+=w_-$ to fit the relative strengths of the weak "vertices" $Y\to N$ and $K\to\pi$. This fit is sensitive to the values of a and b appearing in the (Y,N) vertex and presumably should be predicted by some future theory of weak interactions. We still find, however, that for a range of values for a and b the s-wave part of the Λ decay is essentially the same as the s wave in $\Sigma^+\to p+\pi^0$ decay and thus by choosing $g_{K\Lambda}/g_{K\Sigma}$ to fit w_Λ/w_Σ, almost 100% asymmetry is assured in Λ decay.

One could now use the same approximation to predict $\Sigma^+\to p+\gamma$, and $\Lambda\to n+\gamma$ rates and asymmetries. These rates are proportional to the anomalous moments of the hyperons. However, they are also proportional to the constants a and b which in principle can be determined from the main decay modes.

The remarkable success of the "pole" approximation in predicting the π-decay rate[2] prompts us to believe that the procedure outlined here may provide a reasonable method for determining the magnitudes and signs of the renormalized strong coupling constants, ($g_{\pi\Lambda}, g_{\pi\Sigma}$ and $g_{K\Lambda}/g_{K\Sigma}$, as well as the Ξ coupling constants when experimental data becomes available). A further check on the values of $g_{K\Lambda}$, $g_{K\Sigma}$, and $g_{\pi\Lambda}$ would be provided by the leptonic decays

$$Y\to N+e+\nu,$$
$$\Sigma\to\Lambda+e+\nu,$$

as discussed in references 2 and 3.

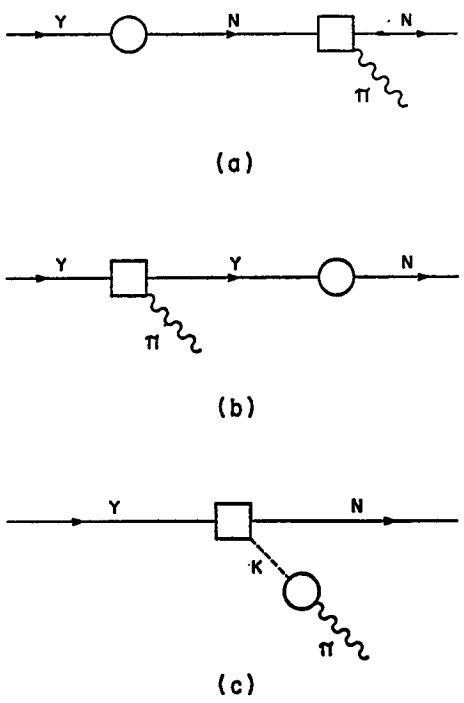


FIG. 1. Graphical representation of the pole terms appearing in hyperon decay into a nucleon and pion. The black box is a purely strong interaction, while the circle involves the weak interaction.

[8] It may be remarked that since T_+ is pure s wave and T_- is pure p wave, (14) is equivalent to $|\alpha_0|=1$ from the $\Delta I=\frac{1}{2}$ rule.

[9] P. T. Matthews and A. Salam, Nuovo cimento 7, 789 (1958).

7.A Nuclear Physics 28 (1961) 529—549; © North-Holland Publishing Co., Amsterdam

CHANGE OF VARIABLES AND EQUIVALENCE THEOREMS IN QUANTUM FIELD THEORIES

S. KAMEFUCHI †, L. O'RAIFEARTAIGH
Dublin Institute for Advanced Studies, Dublin, Ireland
and
ABDUS SALAM
Imperial College of Science and Technology, London, Engeland

Received 20 April 1961

Abstract: In a recent paper, Chisholm has proved the important theorem that the S-matrix in quantum field theories remains unchanged under any point transformation of field operators. In view of some unsatisfactory features in his argument, another proof of this theorem, within the framework of the conventional canonical formalism of field theories, is given. Further, it is pointed out that on the basis of Chisholm's theorem, most of the ordinary equivalence theorems in field theories can be obtained by rather trivial changes of variables in the Lagrangians. The equivalence theorem for the Yang-Mills field is discussed in detail.

1. Introduction

It is well-known in classical mechanics that the Lagrange equations are invariant under any transformation of the variables, which does not involve time derivatives (i.e. any *point* transformation). In quantum field theories, however, it is not yet known whether or not the whole theory remains invariant under similar transformations. In fact, so far the only kind of transformations which have been utilized to any great extent are the unitary (or canonical) transformations.

Recently, Chisholm [1]) has discussed this problem and claimed that any point transformation of variables does not alter the S-matrix, provided that the free field Lagrangian is kept unchanged under the transformation. His argument is based entirely on the asymptotic conditions for the field operators before and after the transformation. It seems to us, however, that the argument in his proof of the above theorem can be justified if and only if the so-called 'strong-convergence' of field operator at $t = \pm\infty$ is assumed. It is well-known from abstract field theory, however, that the strong convergence condition is too strict for interacting fields and has to be replaced by a 'weak convergence' condition. We therefore believe that while Chisholm's theorem

† I.C.I. Research Fellow of University of London at Imperial College, London. Temporarily in Dublin.

is true as a general rule, his proof is not quite complete. In fact, if we apply Chisholm's argument, we are led, in some specially chosen cases, to a contradiction.

In view of this it seems worthwhile to re-examine this problem within the framework of the Heisenberg-Pauli formalism of field theories. We assume in addition, however, the adiabatic switching on and off of interactions, where by interactions we mean any product of *three* or more field operators, possibly all of the same field.

In sects. 2, 3 and 4, we shall try to show that for ordinary field theories, Chisholm's theorem is true, for general point transformations in the case of boson fields, and for a particular class of transformations in the case of fermion fields. Here, by ordinary field theories we mean those theories in which the interaction Lagrangians contain the time derivatives of boson fields at most in a linear form and contain no time derivatives of fermion fields. Although in carrying out our proof we shall have to impose a certain restriction on the transformation of fermion variables (which corresponds roughly to the 'point' nature of the transformations for boson fields), we may expect that the theorem would be true for more general transformations, since we know that it holds for unitary transformations which may in general contain time derivatives. It should be remarked, however, that as far as transformations which do not contain any time derivatives of field operators are concerned the above point transformations are more general than the unitary transformations. In other words, the point transformation and the unitary transformation are quite different things, neither class of transformation being entirely contained within the other.

A second problem which we shall discuss in this paper is that of equivalence theorems. In quantum field theories it has been known for some time that two (or more) alternative expressions for a Lagrangian or a Hamiltonian can lead to an identical final result for physically observable quantities. These relationships, called equivalence theorems, are usually proved by means of unitary transformations or by similar methods, which often involve rather complicated calculations. Once Chisholm's theorem is established, however, one sees that most of the equivalence theorems can be obtained simply by a change of variables in the corresponding Lagrangians.

In sect. 5, we shall re-consider the problem of equivalence theorems from this point of view, and, in particular, give a detailed discussion of the equivalence theorems for the Yang-Mills field with various types of interactions. In this way it is hoped that the connection between the equivalent interactions and the generalized gauge transformation will be made clear.

To keep the main trend of our argument continuous, some of the more mathematical parts of it have been relegated to an Appendix at the end of the paper.

2. Comment on Chisholm's Paper

Consider a Lagrangian for a system of fields $\phi_r(x)$, where $r = 1, 2, \ldots$:

$$\mathscr{L} = \mathscr{L}_0(\phi_1)+\mathscr{L}_0(\phi_2)+\ldots+\mathscr{L}_{\text{int}}(\phi_1\phi_2,\ldots), \tag{2.1}$$

and make the following change of variables:

$$\phi_r \to \phi'_r = \phi_r+F_r(\phi_1, \phi_2, \ldots) \ (r = 1, 2, \ldots), \tag{2.2}$$

where the functions F_r consist of terms which contain *at least two* field operators. Under this change of variables the bi-linear parts of $\mathscr{L}$, that is, the free field Lagrangians, remain unchanged, while the interaction Lagrangian changes, in general. Chisholm's theorem states that the S-matrix remains unchanged under the change of variables (2.2).

In order to prove this, he has shown that the field operators ϕ_r and ϕ'_r have the same asymptotic limits at $t = \pm\infty$, i.e.,

$$\phi^{(\alpha)}_{r\,\text{in(out)}} = \phi'^{(\alpha)}_{r\,\text{in(out)}}, \tag{2.3}$$

or, what is the same thing,

$$F^{(\alpha)}_{r\,\text{in(out)}} = 0. \tag{2.3'}$$

Here, the superscript (α) means the α-component of the operator, defined in the usual way by the scalar product of the operator and a normalizable wave packet $f^{(\alpha)}(x)$. To prove eq. (2.3') he has assumed the relation

$$F_{r\,\text{in(out)}} = \operatorname*{Lim}_{t=\pm\infty} F_r(\phi_1, \phi_2 \ldots) = F_r(\phi_{1\,\text{in(out)}}, \phi_{2\,\text{in(out)}}, \ldots),$$

and then applied the Riemann-Lebesgue theorem to the last expression. Strictly speaking, however, such a procedure is allowed only when all field operators ϕ_r converge *strongly* † to their respective asymptotic limits. It is well-known, however, that the strong convergence is too strict a condition to be satisfied by any interacting fields and has to be replaced by *weak* convergence †. With weak convergence Chisholm's argument is no longer justified. To illustrate this we shall show that in certain cases his argument leads us to a contradiction.

We take, for example, a free neutral scalar field, the Lagrangian of which is given by

$$\mathscr{L} = \mathscr{L}_0 = -\tfrac{1}{2}(\partial_\mu\phi\partial_\mu\phi+\mu^2\phi^2), \tag{2.4}$$

and introduce a new variable ϕ' by

$$\phi = \phi'+f\phi'^2, \tag{2.5}$$

where f is a (coupling) constant. In terms of the new variable $\mathscr{L}$ takes the

† See, for example, ref. [2]) and papers quoted there.

following form †:

$$\mathscr{L} = \mathscr{L}_0(\phi') + \mathscr{L}_{\text{int}}(\phi'),$$
$$\mathscr{L}_{\text{int}}(\phi') = f\phi'^2(\Box - \mu^2)\phi' + \tfrac{1}{2}f^2\phi'^2(\Box - \mu^2)\phi'^2. \tag{2.6}$$

Now, according to Chisholm's argument eq. (2.3) or (2.3') we get

$$f\phi'^{2(\alpha)} \to 0 \qquad (\text{for all } \alpha) \quad \text{as} \quad t \to \pm\infty, \tag{2.7}$$

which means that

$$f\phi'^2 \to 0 \qquad \text{as} \quad t \to \pm\infty. \tag{2.7'}$$

This result cannot be corect for the following reason. According to Chisholm $P_\mu(\phi')$, the energy-momentum vector, should tend to $P_\mu(\phi'_{\text{in(out)}})$, which, in turn, should tend to $P_\mu^{(0)}(\phi'_{\text{in(out)}})$, the energy-momentum vector for the free field $\phi'_{\text{in(out)}}$. However, eq. (2.7') indicates that not only the interaction term, but also the mass term of $\mathscr{L}_0(\phi')$ vanishes. Since the field $\phi'_{\text{in(out)}}$ should have the same mass as $\phi_{\text{in(out)}}$, we arrive at a contradiction.

In this connection it is worthwhile to notice that the appearance of such an inconsistency might also be reduced to the fact that there exists a domain in which the Riemann-Lebesque theorem cannot be properly applied. For example, in Chisholm's equation (2.7) one can choose for the wave functions $f^{(\alpha)}$, a set of plane waves $V^{-\frac{1}{2}}e^{i\mathbf{k}^{(\alpha)}\cdot\mathbf{x}}$, while $f^{(\beta)}, f^{(\gamma)}, \ldots$ are taken as normalizable wave packets. We may apply the Riemann-Lebesque theorem to Chisholm's equation (2.7) with such $f^{(\alpha)}$, $f^{(\beta)}$, $f^{(\gamma)}, \ldots$, but only for $\alpha \neq 0$, or $\mathbf{k}^{(\alpha)} \neq 0$, as can be checked. For $\alpha = 0$ or $k^{(\alpha)} = 0$, the Riemann-Lebesgue theorem does not apply. That is, the above eq. (2.7) or (2.7') is true only for $\alpha \neq 0$. It is true that an argument based on such $f^{(\alpha)}$ is dangerous, because, strictly speaking, the set $f^{(\alpha)}$ is not normalizable. Nevertheless, it is interesting to see that with such a set and taking into account the fact that the Riemann-Lebesgue theorem does not apply for $\alpha = 0$, we can actually get the correct answer $P_\mu(\phi'_{\text{in(out)}}) = P_\mu^{(0)}(\phi'_{\text{in(out)}})$. This is because $P_\mu^{(0)}(\phi'_{\text{in(out)}})$ consists *only* of $P_\mu(\phi'_{\text{in(out)}})^{(\alpha)}$ with $\alpha = 0$ (all the components with $\alpha \neq 0$ vanish).

In view of the rather delicate nature of the asymptotic condition in abstract field theory (exemplified by our above results) we shall try, in the following section, to construct a proof of Chisholm's theorem, which lies within the older, Hamiltonian, formalism.

3. Invariance under Point Transformations in the Heisenberg-Pauli Formalism

3.1. PRELIMINARY CONSIDERATION

In this section we shall work exclusively in the Heisenberg representation Our aim will be to establish the invariance of the Heisenberg-Pauli formalism [4])

† To deal with such a system as eq. (2.6) with high derivatives one should have recourse, for example, to the general method proposed by Umezawa and Takahashi [3]).

under point transformations. To illustrate the idea involved we shall indicate first how this invariance is established in the case of classical fields. Let $\phi_r(x)$, $r = 1, 2, \ldots n$, be a set of n classical fields, whose motions are determined by a Lagrangian density $\mathscr{L}(\phi_r, \dot{\phi}_r)$ involving only the ϕ_r and their first time derivatives. The equations of motion are

$$\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}_r} = \frac{\delta \mathscr{L}}{\delta \phi_r}. \tag{3.1}$$

We now make a point transformation to new variables

$$\phi'_s(x) = \phi'_s(\phi_r(x)), \qquad \phi_r(x) = \phi_r(\phi'_s(x)), \tag{3.2}$$

i.e., any transformation (whose Jacobian is not zero) such that the ϕ' involve the ϕ but not the $\dot{\phi}$. It is then required to show that

$$\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s} = \frac{\delta \mathscr{L}}{\delta \phi'_s}, \tag{3.3}$$

$$T_{\mu\nu} \equiv \sum_r \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \phi_{r,\nu} = T'_{\mu\nu} \equiv \sum_s \frac{\partial \mathscr{L}}{\partial \phi'_{s,\mu}} \phi'_{s,\nu}. \tag{3.4}$$

The proof is fairly simple:

$$\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s} = \frac{\partial}{\partial t}\sum_r \frac{\partial \mathscr{L}}{\partial \dot{\phi}_r}\frac{\partial \dot{\phi}_r}{\partial \dot{\phi}'_s} = \sum_r \left(\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}_r}\right)\frac{\partial \phi_r}{\partial \dot{\phi}'_s} + \sum_r \frac{\partial \mathscr{L}}{\partial \dot{\phi}_r}\frac{\partial}{\partial t}\frac{\partial \phi_r}{\partial \dot{\phi}'_s}\Big)$$

$$= \sum_r \frac{\delta \mathscr{L}}{\delta \phi_r}\frac{\partial \phi_r}{\partial \dot{\phi}'_s} + \sum_r \frac{\partial \mathscr{L}}{\partial \dot{\phi}_r}\frac{\partial \dot{\phi}_r}{\partial \dot{\phi}'_s} = \frac{\delta \mathscr{L}}{\delta \phi'_s},$$

as required. Note that we have used the relations

$$\frac{\partial \dot{\phi}_r}{\partial \dot{\phi}'_s} = \frac{\partial \phi_r}{\partial \phi'_s} \qquad \text{and} \qquad \frac{\partial}{\partial t}\frac{\partial \phi_r}{\partial \phi'_s} = \frac{\partial \dot{\phi}_r}{\partial \phi'_s}, \tag{3.5}$$

which can easily be verified from (3.2). Similarly, we have

$$T'_{\mu\nu} \equiv \sum_s \frac{\partial \mathscr{L}}{\partial \phi'_{s,\mu}} \phi'_{s,\nu} = \sum_s \sum_r \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}}\frac{\partial \phi_{r,\mu}}{\partial \phi'_{s,\mu}} \phi'_{s,\nu}$$

$$= \sum_r \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \sum_s \frac{\partial \phi_r}{\partial \phi'_s} \phi'_{s,\nu} = \sum_r \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \phi_{r,\nu} \equiv T_{\mu\nu}.$$

In this way the invariance of the theory under point transformations is established in the classical theory. As is seen, what is meant here by 'invariance' is this: if we start with one set of variables ϕ and use the Lagrangian formalism we arrive at the same result as if we first changed the variables ϕ to new variables ϕ' in the Lagrangian, then applied the Lagrangian formalism to the

ϕ', and afterwards expressed the results (field equations, etc.) in terms of the original ϕ. In other words, the field equations, etc., for the ϕ and the ϕ' are different in form but equal in content.

Let us now consider the question of invariance in this sense in quantum field theory. In this case we find that the situation is not so simple. If one attempts a proof analogous to the above one encounters two difficulties. The first is that due to the non-commutativity of the ϕ_r and $\dot{\phi}_r$, the direct proofs just given for the classical case break down (for example, it is no longer true that $\partial\mathscr{L}/\partial\dot{\phi}'_s = \sum_r(\partial\mathscr{L}/\partial\dot{\phi}_r)\ (\partial\dot{\phi}_r/\partial\dot{\phi}'_s)$). The second difficulty is that we must establish, in addition to (3.3) and (3.4), the further invariance

$$\left[\frac{\partial\mathscr{L}}{\partial\dot{\phi}_r}, \phi_s\right] = \left[\frac{\partial\mathscr{L}}{\partial\dot{\phi}'_r}, \phi'_s\right] = \Delta\delta_{rs}, \qquad \text{etc.,} \tag{3.6}$$

and

$$[P_\mu, \phi_r] = i\phi_{r,\mu}, \qquad [P'_\mu, \phi'_s] = i\phi'_{s,\mu}, \tag{3.6}$$

where Δ is a (singular) c-number.

It is clear, therefore, that the problem in quantum field theory is more complex than in classical theory. We consider it for boson fields and fermion fields separately.

3.2. BOSON FIELDS

First, we replace the general point transformation (3.2) of the classical theory by transformations

$$\phi_r = \phi_r(\phi'_s), \qquad \phi'_s = \phi'_s(\phi_r), \tag{3.7}$$

such that ϕ_r can be expressed as a power series in terms of the ϕ'_s and vice versa. (This is a familiar restriction on functions in quantum mechanics as well as in quantum field theory. An example of such a transformation is

$$\phi = \phi' - \phi'^2 + \ldots + (-1)^{n-1}\phi'^n + \ldots, \qquad \phi' = \phi + \phi^2 + \ldots + \phi^n + \ldots).$$

We define point transformations of boson variables to be transformations of this kind only.

Due to the commutativity of the ϕ_r a first result of eqs. (3.7) is that for $\phi''_k = \phi''_k(\phi'_s)$ we get

$$\frac{\partial\phi_r}{\partial\phi''_k} = \sum_s \frac{\partial\phi_r}{\partial\phi'_s}\frac{\partial\phi'_s}{\partial\phi''_k}. \tag{3.8}$$

We now consider the general definition of derivatives with respect to field variables. Let $F(\phi_r, \dot{\phi}_r)$ be any function of the ϕ_r, $\dot{\phi}_r$ which can be expressed as a (possibly infinite) sum of simple products of the quantities, e.g. $F = \phi + (\phi\dot{\phi} + \dot{\phi}\phi) + \phi\dot{\phi}^2$. Since $[\phi_r, \dot{\phi}_r] \neq 0$, the question is, what do we mean by

$\partial F/\partial\phi_r$ and $\partial F/\partial\dot{\phi}_r$? We define these quantities in three steps, in agreement with the original definition of Heisenberg and Pauli [4]):

Step 1:

$$\frac{\partial\phi_r}{\partial\phi_s}=\frac{\partial\dot{\phi}_r}{\partial\dot{\phi}_s}=\delta_{rs}, \qquad \frac{\partial\phi_r}{\partial\dot{\phi}_s}=\frac{\partial\dot{\phi}_r}{\partial\phi_s}=0. \tag{3.9}$$

Step 2:

Let X or Y or Z denote ϕ_r or $\dot{\phi}_r$ ($r = 1, 2, \ldots, n$). Then we put

$$\frac{\partial}{\partial Z}XY=\frac{\partial X}{\partial Z}Y+X\frac{\partial Y}{\partial Z}, \tag{3.10}$$

with the generalization, for $X_i = \phi_r$ or $\dot{\phi}_r$ ($r = 1, 2, \ldots, n$),

$$\frac{\partial}{\partial Z}\prod_{i=1}^{N}X_i=\sum_{j=1}^{N}\prod_{i=1}^{j-1}X_i\frac{\partial X_j}{\partial Z}\prod_{i=j+1}^{N}X_i. \tag{3.11}$$

Step 3:

If F_m is a simple product as in Step 2 and $F = \sum_{m=1}^{M} F_m$, we put

$$\frac{\partial}{\partial Z}F=\sum_{m=1}^{M}\frac{\partial F_m}{\partial Z}. \tag{3.12}$$

In this way the derivatives with respect to ϕ_r, $\dot{\phi}_r$ are unambiguously defined for functions $F(\phi_r, \dot{\phi}_r)$ of the kind mentioned (these are the only ones which we shall consider).

We are now in a position to examine the invariance of the Heisenberg-Pauli formalism under point transformations. Before proceeding to the investigation, however, it is useful to state two important theorems which follow directly form our definitions and which are proved in the Appendix. They are

$$\frac{\partial}{\partial\phi'_s}\frac{\partial}{\partial t}\phi_r(\phi'_s)=\frac{\partial}{\partial\phi'_s}\dot{\phi}_r(\phi'_s) \qquad \text{(Theorem I)},$$

$$\frac{\partial}{\partial t}\frac{\partial}{\partial\phi'_s}\phi_r=\frac{\partial}{\partial\phi'_s}\dot{\phi}_r \qquad \text{(Theorem II)}.$$

It will be noticed that these are just the relations (3.5) which played such a vital role in the proof of invariance for the classical fields. We proceed now to the proof of invariance.

Let ϕ_r ($r = 1, 2, \ldots, n$) be a set of n boson fields with Lagrangian density

$$\mathscr{L}=\sum_{r=1}^{n}\tfrac{1}{2}\dot{\phi}_r^2+G(\phi_r). \tag{3.13}$$

where G is any function (expressible as a power series) of the ϕ_r. The Heisenberg

Pauli formalism

$$\left[\frac{\partial \mathscr{L}}{\partial \dot{\phi}_r}, \phi_s\right] = \Delta\delta_{rs}, \qquad \frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}_r} = \frac{\delta \mathscr{L}}{\delta \phi_r},$$
$$T_{\mu\nu} = \tfrac{1}{2}\sum_q \left\{\phi_{q,\nu}\frac{\partial \mathscr{L}}{\partial \phi_{q,\mu}} + \frac{\partial \mathscr{L}}{\partial \phi_{q,\mu}}\phi_{q,\nu}\right\} - \delta_{\mu\nu}\mathscr{L}, \qquad [P_\mu, \phi_r] = i\phi_{r,\mu}, \tag{3.14}$$

is equivalent in this case to

$$[\dot{\phi}_r, \phi_s] = \Delta\delta_{rs}, \qquad \ddot{\phi}_r = \frac{\delta G}{\delta \phi_r}, \qquad T_{44} = \sum_q \tfrac{1}{2}\dot{\phi}_q^2 - G(\phi_r), \text{ etc.} \tag{3.15}$$

Now let $\phi'_s = \phi'_s(\phi_r)$ (point transformation); then one has

$$\left[\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s}, \phi'_r\right] = \Delta\delta_{rs}, \qquad \frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s} = \frac{\delta \mathscr{L}}{\delta \phi'_s}, \qquad T_{\mu\nu} = T'_{\mu\nu}. \tag{3.16}$$

We prove this in the following way. By definition and theorem I we have

$$\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s} = \tfrac{1}{2}\sum_q \left(\dot{\phi}_q \frac{\partial \dot{\phi}_q}{\partial \dot{\phi}'_s} + \frac{\partial \dot{\phi}_q}{\partial \dot{\phi}'_s}\dot{\phi}_q\right) = \tfrac{1}{2}\sum_q \left(\dot{\phi}_q \frac{\partial \phi_q}{\partial \phi'_s} + \frac{\partial \phi_q}{\partial \phi'_s}\dot{\phi}_q\right),$$

and therefore, using eqs. (3.15) and (3.8)

$$\left[\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s}, \phi'_r\right] = \tfrac{1}{2}\sum_q \left\{[\dot{\phi}_q, \phi'_r]\frac{\partial \phi_q}{\partial \phi'_s} + \frac{\partial \phi_q}{\partial \phi'_s}[\dot{\phi}_q, \phi'_r]\right\}$$
$$= \tfrac{1}{2}\sum_q \left\{\Delta \frac{\partial \phi'_r}{\partial \phi_q}\frac{\partial \phi_q}{\partial \phi'_s} + \Delta \frac{\partial \phi_q}{\partial \phi'_s}\frac{\partial \phi'_r}{\partial \phi_q}\right\} = \Delta \frac{\partial \phi'_r}{\partial \phi'_s} = \Delta\delta_{rs}.$$

Further, using eq. (3.15) and theorem II, we get

$$\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{\partial \dot{\phi}'_s} = \tfrac{1}{2}\sum_q \left\{\ddot{\phi}_q \frac{\partial \phi_q}{\partial \phi'_s} + \frac{\partial \phi_q}{\partial \phi'_s}\ddot{\phi}_q + \dot{\phi}_q\left(\frac{\partial}{\partial t}\frac{\partial \phi_q}{\partial \phi'_s}\right) + \left(\frac{\partial}{\partial t}\frac{\partial \phi_q}{\partial \phi'_s}\right)\dot{\phi}_q\right\}$$
$$= \tfrac{1}{2}\sum_q \left\{\frac{\delta G}{\delta \phi_q}\frac{\partial \phi_q}{\partial \phi'_s} + \frac{\partial \phi_q}{\partial \phi'_s}\frac{\delta G}{\delta \phi_q} + \dot{\phi}_q \frac{\partial \dot{\phi}_q}{\partial \phi'_s} + \frac{\partial \dot{\phi}_q}{\partial \phi'_s}\dot{\phi}_q\right\}$$

and finally, by a slight extension of eq. (3.8) to include functional derivatives,

$$= \frac{\delta G}{\delta \phi'_s} + \frac{\partial}{\partial \phi'_s}\sum_q \tfrac{1}{2}\dot{\phi}_q^2 = \frac{\delta \mathscr{L}}{\delta \phi'_s}.$$

We have still to show that $T_{\mu\nu} = T'_{\mu\nu}$; proof has been relegated to the Appendix as it involves a slightly more general definition of $T_{\mu\nu}$ than the usual one, and this makes it rather lengthy. Finally, the invariance of $[P_\mu, \phi_r] = i\phi_{r,\mu}$ follows immediately from that of $T_{\mu\nu}$ and of the commutation rules.

We now discuss briefly some points in connection with the above proof.

First, we note that the invariance has been shown only for $\partial/\partial\phi_r$ and $\partial/\partial\dot{\phi}_r$ as defined above (Steps 1, 2, 3). Secondly we started from a very special Lagrangian, involving the $\dot{\phi}_r$ only separately, quadratically and with numerical coefficients. As a matter of fact, the proof also holds good (as can easily be checked) if G involves the $\dot{\phi}_r$ in a linear way. However, it is not so certain (and in fact it seems to us at present to be unlikely) that the proof holds if G involves either $\dot{\phi}_r{}^2 f(\phi_r)$ or $\ddot{\phi}_r$. Thus the above proof is, so to speak, irreversible: We start with a standard Lagrangian (3.13), build another more general one (e.g. involving terms like $\dot{\phi}_r{}^2 f(\phi_r)$) by means of a point transformation, and then show the equivalence of the two. But, given some general Lagrangian (involving $\dot{\phi}_r{}^2 f(\phi_r)$, say) we have not shown that this is invariant under an arbitrary point transformation (much less have we shown that it can be reduced to the standard form).

In this connection we also observe the following point. For a Lagrangian transformed by an arbitrary point transformation, the transition from Lagrangian to Hamiltonian is not so simple in general because of the presence of terms bilinear in time derivatives. To express $\dot{\phi}'_r$ in terms of canonical variables we may have to have recourse to power series expansion in coupling constants. As for the calculation of the S-matrix in such a case a detailed discussion has been given in ref. [3]).

3.3. FERMION FIELDS

Much of the formalism for fermion fields is the same as in the boson case, and we therefore relegate the details to the Appendix and present here only the new features. The first new feature is the definition of 'point' transformations. Let ψ be a spinor. Then a point transformation of ψ is

$$\psi = \psi' F(\psi'), \tag{3.17}$$

where F is a function only of bilinear forms of ψ' and does not contain $\bar{\psi}'$. (this kind of term does not generally conserve the particle number). The property of F of being bilinear in ψ' is not a serious restriction as it is necessary in any case in order that ψ' transform like ψ under a Lorentz transformation. The restriction that F cannot contain $\bar{\psi}'$ is, on the contrary, a strong one and is analogous to the condition that ϕ'_s does not contain $\dot{\phi}_r$ in the boson case. Thus, neither in the boson nor in the fermion case does the class of point transformations include the general unitary transformation.

With this definition of point transformations, and with a standard Lagrangian

$$\mathscr{L} = -\tfrac{1}{2}\bar{\psi}(\gamma\overrightarrow{\partial}+m)\psi+\tfrac{1}{2}\bar{\psi}(\gamma\overleftarrow{\partial}-m)4+g\bar{\psi}(\gamma\text{'s; boson operators})\psi,$$

the proof of invariance proceeds analogously to the boson case. There is a slight change of signs in the definition of $\partial/\partial\psi_\alpha$, etc. due to the anti-commuta-

tion, but eq. (3.8) and theorems I and II hold as before. The transformation (3.17) can also be generalized to include boson variables in F.

To summarize, therefore, the general result of this section is that, with the definitions given, the Heisenberg-Pauli formalism is invariant under the point transformation

$$\psi = \psi' F(\psi', \phi'), \qquad \phi = \phi(\phi'), \tag{3.18}$$

F being bilinear in ψ'.

4. Proof of Equality of S-Matrices

In the preceding section we have shown that the Lagrangian-Hamiltonian formalism is invariant under the point transformation (3.18). This means that it does not matter which set of variables ϕ or ϕ' one may choose in describing a given physical system; it is merely a matter of convenience. Of course, the physical meanings to be attached to these variables will in general vary according to the choice of variables.

In quantum field theory, we usually choose the variables in such a way that the free field operators obey the equations of motion for a simple harmonic oscillator. This allows us to make the usual particle interpretation of fields. Thus, if we restrict ourselves to those variables which satisfy this requirement, the change of variables is to be limited to the following type:

$$\phi'_r(x) = \phi_r(x) + g_r F_r(g_s; \phi_s(x)), \qquad (r, s = 1, 2, \ldots, n), \tag{4.1}$$

where the F_r are functions of the ϕ, each being a sum of terms which are products of at least two field operators. For dimensional reasons, constants g_s are introduced here, which may be regarded as coupling constants, for in the Lagrangian they appear only in 'interaction' terms, by which we mean terms containing *three* or more field operators (possibly all of the same field).

Now, as was proved in eq. (3.16), we have for the total Hamiltonian (in the Schrödinger representation)

$$H = H_0(\phi) + H_{\text{int}}(\phi) = H' = H_0(\phi') + H'_{\text{int}}(\phi'). \tag{4.2}$$

Let us consider two interaction representations, in one of which H_{int}, and in the other of which H'_{int} appears in the Schrödinger equation, the respective state vectors being $\Psi(t)$ and $\Psi'(t)$. Since $\Psi(t)$ and $\Psi'(t)$ are connected with the common Schrödinger (or Heisenberg) representation by unitary transformations, they are mutually related by a unitary transformation $U(t)$,

$$\Psi(t) = U(t)\Psi'(t), \tag{4.3}$$

with, for example,

$$U(t) = e^{-iH_0(\phi_{t=0})t} e^{iH_0(\phi'_{t=0})t}.$$

It is important to notice here that since these two representations differ only by the terms depending on the interaction, they should coincide at $t = \pm\infty$, when both H_{int} and H'_{int} are switched off under the adiabatic hypothesis which, as we have said, we assume for all interactions (products of *three* or more operators) as in the ordinary theory. Thus, we have

$$U(t) \to 1 \qquad \text{as} \qquad t \to \pm\infty. \tag{4.4}$$

This can also be seen by rewriting $U(t)$ in the form

$$U(t) = \exp\{i[H'_{\text{int}} - H_{\text{int}}]t + \tfrac{1}{2}[H'_{\text{int}}, H_{\text{int}}]t^2 + \ldots\}.$$

It should be remarked here that the above requirement which implies "inter alia" that $\phi^n \to 0$, $n \geqq 3$) is weaker than Chisholm's asymptotic condition which, as we have seen, implied $\phi^2 \to 0$ (so that $\phi = \phi' + f\phi'^2 \to \phi'$) and so led to a contradiction with the physical requirement that the mass term in the Lagrangian remain non-zero for $t \to \pm\infty$ †.

Corresponding to the above two interaction representations we get two different expressions for the S-matrix, S and S', which can be obtained by taking H_{int} and H'_{int}, respectively, as interaction Hamiltonian. From (4.3) S and S' are connected by

$$S' = U^{-1}(+\infty)SU(-\infty),$$

and using (4.4) we now get

$$S' = S. \tag{4.5}$$

This means that two different interaction Hamiltonians $H_{\text{int}}(\phi)$ and $H'_{\text{int}}(\phi')$ lead to identical S-matrices, and this is exactly Chisholm's result (apart from the restrictions imposed on the transformations of fermion variables in our case). Finally, let us remark that while the Heisenberg operators ϕ and ϕ' are not necessarily connected by a unitary transformation (in this sense our class of point transformations is wider than the class of unitary transformations) the operators ϕ and ϕ' in the respective interaction representations are thus connected, because they satisfy the same commutation relations.

5. Equivalence Theorems

In quantum field theories there exist several kinds of theorems, called equivalence theorems, which state that two alternative expressions for the interaction Lagrangian or Hamiltonian lead to identical final answers for phys-

† Of course, our assumption that terms such as $\phi^n \to 0$, $n \geqq 3$ may not be quite free from objection from a strict mathematical point of view either, but we feel safer to make this assumption, since (in contrast to the assumption $\phi^2 \to 0$) it does not seem to lead to any physical contradiction. Note also that in our case, although $H_0(\phi) = H_0(\phi')$ at $t = \pm\infty$, this does not necessarily imply $\phi = \phi'$.

ically observable quantities. These theorems have so far been proved by means of unitary transformations or by similar methods, which are not in general simple.

However, when viewed from the standpoint of the theorem established above, the equivalence theorems are in most cases an immediate coesequence of simple changes of variables in the original Lagrangians. Almost trivial examples of this are the proofs of gauge invariance of quantum electrodynamics [5]), the equivalence theorems in meson theories discussed by Nelson, Dyson and Case [6]), and the equivalence theorem in neutral vector meson theory discussed recently by two of the present authors [7]).

To illustrate, let us take as an example the Lagrangian for the π-meson–nucleon system given by

$$\mathscr{L} = -\bar{\psi}(\gamma\partial+m)\psi-\tfrac{1}{2}(\partial_\mu\boldsymbol{\phi}\cdot\partial_\mu\boldsymbol{\phi}+\mu^2\boldsymbol{\phi}\cdot\boldsymbol{\phi})+ig\bar{\psi}\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi}\psi, \tag{5.1}$$

and make the following change of variables:

$$\psi\to\psi'=\psi\exp\left[-\frac{ig}{2m}\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi}\right], \qquad \boldsymbol{\phi}\to\boldsymbol{\phi}'=\boldsymbol{\phi}, \tag{5.2}$$

which satisfy the conditions (3.18) and (4.1) for the F in our invariance theorem. Substituting (5.2) into (5.1) and omitting the prime one gets

$$\begin{aligned}\mathscr{L} = {} & -\bar{\psi}(\gamma\partial+m)\psi-\tfrac{1}{2}(\partial_\mu\boldsymbol{\phi}\cdot\partial_\mu\boldsymbol{\phi}+\mu^2\boldsymbol{\phi}\cdot\boldsymbol{\phi}) \\ & -m\bar{\psi}\left(\exp\left[\frac{ig}{m}\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi}\right]-1-\frac{ig}{m}\gamma_5\boldsymbol{\tau}\cdot\boldsymbol{\phi}\right)\psi+\frac{ig}{m}\bar{\psi}\gamma_5\gamma_\mu\left[f\left(-g,\frac{\gamma_5\boldsymbol{\phi}}{m}\right)\partial_\mu\boldsymbol{\phi}\right]\cdot\boldsymbol{\tau}\psi,\end{aligned} \tag{5.3}$$

where $f(g, \mathbf{a})$ is the operator in iso-space introduced in a previous paper [8]) and given by

$$f(g,\mathbf{a})=\frac{\exp[g\alpha(\mathbf{a})]-1}{g\alpha(\mathbf{a})}, \tag{5.4}$$

with $\alpha(\mathbf{a})$ such that $\alpha(\mathbf{a})\mathbf{b}=\mathbf{b}\wedge\mathbf{a}$, where $\mathbf{b}$ is an arbitrary iso-vector. Eq. (5.3) is essentially the result obtained by Drell and Henly [9]), and when expanded in powers of g, it gives the familiar expression of Nelson, Dyson and Case.

The equivalence theorem in neutral vector meson theory with partial conservation law can be proved in a similar manner.

6. Equivalence Theorem for the Yang-Mills Field

6.1. THE FREE YANG-MILLS FIELD

The Lagrangian is given by †

$$\mathscr{L}_{\text{Y.M.}}=\mathscr{L}_{\text{Y.M.0}}(\mathbf{u}_\mu)+\mathscr{L}_{\text{Y.M.int}}(\mathbf{u}_\mu), \tag{6.1}$$

† In this section we use the same notation as in ref. [8]).

where

$$\begin{aligned}\mathscr{L}_{\text{Y.M.}\,0}\ (u_\mu) &= -\tfrac{1}{4}\mathbf{u}_{\mu\nu}\cdot\mathbf{u}_{\mu\nu}+\tfrac{1}{2}\kappa^2\mathbf{u}_\mu\cdot\mathbf{u}_\mu,\\ \mathscr{L}_{\text{Y.M. int}}(u_\mu) &= -\tfrac{1}{2}g\mathbf{u}_{\mu\nu}\cdot(\mathbf{u}_\mu\wedge\mathbf{u}_\nu)-\tfrac{1}{4}g^2(\mathbf{u}_\mu\wedge\mathbf{u}_\nu)\cdot(\mathbf{u}_\mu\wedge\mathbf{u}_\nu).\end{aligned} \tag{6.1'}$$

Eq. (6.1) is invariant except for the mass term, under the following transformation:

$$\mathbf{u}_\mu\to\mathbf{u}'_\mu = S(g,\Lambda)\mathbf{u}_\mu+f(g,\Lambda)\partial_\mu\Lambda, \tag{6.2}$$

where $S(g,\Lambda)$ is defined by

$$S(g,\Lambda) = \exp[g\alpha(\Lambda)] \tag{6.2'}$$

and $f(g,\Lambda)$ is the same operator as that defined in eq. (5.4) †.

In terms of the Stueckelberg variables eq. (6.1) becomes

$$\mathscr{L}_{\text{Y.M.}} = \mathscr{L}_{\text{Y.M.}\,0}(\mathbf{A}_\mu,\mathbf{B})+\mathscr{L}_{\text{Y.M. int}}(\mathbf{U}_\mu), \tag{6.3}$$

where

$$\begin{aligned}\mathscr{L}_{\text{Y.M.}\,0}(\mathbf{A}_\mu,\mathbf{B}) = &-\tfrac{1}{2}\partial_\mu\mathbf{A}_\nu\cdot\partial_\mu\mathbf{A}_\nu-\tfrac{1}{2}\kappa^2\mathbf{A}_\mu\mathbf{A}_\mu\\ &-\tfrac{1}{2}\partial_\mu\mathbf{B}\cdot\partial_\mu\mathbf{B}-\tfrac{1}{2}\kappa^2\mathbf{B}\cdot\mathbf{B}+\tfrac{1}{2}(\partial_\mu\mathbf{A}_\mu+\kappa\mathbf{B})^2,\end{aligned} \tag{6.3'}$$

$$\mathbf{U}_\mu = \mathbf{A}_\mu+\frac{1}{\kappa}\partial_\mu\mathbf{B}\equiv\mathbf{A}_\mu+\partial_\mu\mathbf{b}.$$

The transformation (6.2) now takes the following form:

$$\begin{aligned}\mathbf{A}_\mu\to\mathbf{A}'_\mu &= S(g,\Lambda)\mathbf{A}_\mu+[f(g,\Lambda)-1]\partial_\mu\Lambda,\\ \partial_\mu\mathbf{B}\to\partial_\mu\mathbf{B}' &= S(g,\Lambda)\partial_\mu\mathbf{B}+\kappa\partial_\mu\Lambda.\end{aligned} \tag{6.4}$$

Now, we can show that when making a change of variables, such as

$$\mathbf{A}_\mu\to\mathbf{A}'_\mu = S(g,\mathbf{b})\mathbf{A}_\mu+[f(g,\mathbf{b})-1]\partial_\mu\mathbf{b},\qquad \mathbf{B}\to\mathbf{B}'=\mathbf{B}, \tag{6.5}$$

or

$$\mathbf{U}_\mu\to\mathbf{U}'_\mu = S(g,\mathbf{b})\mathbf{A}_\mu+f(g,\mathbf{b})\partial_\mu\mathbf{b}\equiv\mathbf{W}_\mu, \tag{6.5'}$$

the Lagrangian is transformed into

$$\begin{aligned}\mathscr{L}'_{\text{Y.M.}} &= \mathscr{L}_{\text{Y.M.}\,0}(\mathbf{A}'_\mu,\mathbf{B}')+\mathscr{L}_{\text{Y.M. int}}(\mathbf{U}'_\mu)\\ &= \mathscr{L}_{\text{Y.M.}\,0}(\mathbf{A}_\mu,\mathbf{B})+\mathscr{L}_{\text{Y.M. int}}(\mathbf{A}_\mu)+\tfrac{1}{2}\kappa^2[\mathbf{U}_\mu\cdot\mathbf{U}_\mu-\mathbf{W}_\mu\cdot\mathbf{W}_\mu].\end{aligned} \tag{6.6}$$

The reason why the last term in eq. (6.6) appears can be understood in the following way. Since the change of variables (6.5'), $\mathbf{U}'_\mu\to\mathbf{A}_\mu$, is essentially of the same form as (6.2), $\mathbf{u}'_\mu\to\mathbf{u}_\mu$ (by replacing $\mathbf{u}_\mu$ and Λ in (6.2) by $\mathbf{A}_\mu$ and $\mathbf{b}$, respectively, one can get (6.5')), the non-gauge invariant term $\frac{1}{2}\kappa^2\mathbf{U}'_\mu\mathbf{U}'_\mu$ gives rise to $\frac{1}{2}\kappa^2\mathbf{W}_\mu\mathbf{W}_\mu$.

Eq. (6.6) is exactly the expression which was obtained before by means of a unitary transformation [8]). Thus, we can justify, a posteriori, the above change

† For detailed properties of the operators S and f, see eq. (26) of ref. [8]).

of variables (6.5) or (6.5'), although the invariance of the theory under this transformation is *not* guaranteed by our theorem because of the presence of the time derivative of **B** in the transformation of $\mathbf{A}_4$. This result seems to suggest that our theorem could be generalized to include such cases as well.

Generalizing the argument one can make the following statement: Suppose that a Lagrangian $\mathscr{L}$ is of the form

$$\mathscr{L} = \mathscr{L}_{\text{Y.M.}}(\mathbf{u}_\mu)+\mathscr{L}_1(\mathbf{u}_\mu, \phi)+\mathscr{L}_2(\mathbf{u}_\mu, \phi), \tag{6.7}$$

where the ϕ are field operators interacting with $\mathbf{u}_\mu$. The quantities $\mathscr{L}_1$ and $\mathscr{L}_2$ are invariant and non-invariant parts of $\mathscr{L}$ under the gauge transformation (6.2), combined with some appropriately defined transformations for the ϕ, which are needed to make $\mathscr{L}_{\text{Y.M.}}$ (mass-term) and $\mathscr{L}_1$ invariant. Furthermore. we shall assume that the latter gauge transformation satisfies the condition for point transformations of our theorem. Then, we can say that $\mathscr{L}$ is equivalent to the following expression:

$$\begin{aligned}\mathscr{L} = \mathscr{L}_{\text{Y.M.0}}(\mathbf{A}_\mu, \mathbf{B})+\mathscr{L}_{\text{Y.M. int}}(\mathbf{A}_\mu)+\tfrac{1}{2}\kappa^2[\mathbf{U}_\mu\cdot\mathbf{U}_\mu-\mathbf{W}_\mu\cdot\mathbf{W}_\mu]\\ +\mathscr{L}_1(\mathbf{A}_\mu, \phi)+\mathscr{L}_2(\mathbf{W}_\mu, \phi'),\end{aligned} \tag{6.7'}$$

where ϕ' is obtained from ϕ by making the gauge transformation with the function **b** (not with Λ). That eq. (6.7') follows from (6.7) is due to the fact that the transformation to 'eliminate' the **b**–field is of the same form as the gauge transformation. In the following we shall consider some concrete examples.

6.2. GÜRSEY'S LAGRANGIAN

Following Gürsey [11]) we can write †

$$\mathscr{L}_{\text{G}} = -\tfrac{1}{4}\mathbf{u}_{\mu\nu}\cdot\mathbf{u}_{\mu\nu}-\tfrac{1}{2}\kappa^2\left[\mathbf{u}_\mu+\frac{1}{\kappa}f(-g, \phi/\kappa)\partial_\mu\phi\right]^2+\mathscr{L}_{\text{Y.M. int}}(\mathbf{u}_\mu). \tag{6.8}$$

One can easily check that (6.8) is *invariant* under the following transformation:

$$\begin{aligned}&\mathbf{u}_\mu \to \mathbf{u}'_\mu = S(g, \Lambda)\mathbf{u}_\mu+f(g, \Lambda)\partial_\mu\Lambda,\\ &\exp[ig(\boldsymbol{\tau}\cdot\phi)/2\kappa] \to \exp[ig(\boldsymbol{\tau}\cdot\phi')/2\kappa] = \exp[ig(\boldsymbol{\tau}\cdot\phi)/2\kappa]\exp[-ig(\boldsymbol{\tau}\cdot\Lambda)/2],\\ &f(-g, \phi/\kappa)\partial_\mu\phi \to f(-g, \phi'/\kappa)\partial_\mu\phi' = S(g, \Lambda)f(-g, \phi/\kappa)\partial_\mu\phi-\kappa f(g, \Lambda)\partial_\mu\Lambda,\end{aligned} \tag{6.9}$$

which is a point transformation satisfying our condition. In this case we

† The original expression for eq. (6.8) given by Gürsey is

$$-\tfrac{1}{2}\kappa^2[\mathbf{u}_\mu-ig^{-1}\,\mathrm{Tr}\,\boldsymbol{\tau}\,e^{-ig(\boldsymbol{\tau}\cdot\phi)/2\kappa}\,\partial_\mu\,e^{ig(\boldsymbol{\tau}\cdot\phi)/2\kappa}]^2+\ldots$$

have $\mathscr{L}_2 = 0$. Thus, one gets the following equivalent Lagrangian:

$$\mathscr{L}_G = \mathscr{L}_{Y.M.0}(\mathbf{A}_\mu, \mathbf{B}) + \mathscr{L}_{Y.M.\,int}(\mathbf{A}_\mu) + \tfrac{1}{2}\kappa^2 \mathbf{U}_\mu \cdot \mathbf{U}_\mu - \tfrac{1}{2}\kappa^2 \left[\mathbf{A}_\mu + \frac{1}{\kappa} f(-g, \phi|k)\partial_\mu \phi\right]^2. \quad (6.8')$$

6.3. GELL-MANN-LÉVY MODEL

For the Gell-Mann-Lévy model (in interaction with the Yang–Mills field[12]) we have

$$\begin{aligned}\mathscr{L}_{G.L.} = \mathscr{L}_{Y.M.} &- \bar{\psi}\{\gamma_\mu[\partial_\mu - \tfrac{1}{4}ig(1+\gamma_5)(\tau \cdot \mathbf{u}_\mu)] - h[\sigma + i(\tau \cdot \pi)\gamma_5]\}\psi \\ &- \tfrac{1}{2}[\partial_\mu \pi - \tfrac{1}{2}g\pi \wedge \mathbf{u}_\mu + \tfrac{1}{2}g\sigma \mathbf{u}_\mu]^2 - \tfrac{1}{2}[\partial_\mu \sigma - \tfrac{1}{2}g\pi \cdot \mathbf{u}_\mu]^2 \quad (6.10) \\ &- \tfrac{1}{2}\mu^2[\pi^2 + \sigma^2] - \lambda\left[\pi^2 + \sigma^2 - \frac{1}{4f}\right]^2 - \frac{\mu^2}{2f}\sigma.\end{aligned}$$

The above Lagrangian is invariant, except for the last term and the mass term of $\mathscr{L}_{Y.M.}$, under the following transformation:

$$\begin{aligned}\mathbf{u}_\mu &\to \mathbf{u}'_\mu = S(g, \Lambda)\mathbf{u}_\mu + f(g, \Lambda)\partial_\mu \Lambda, \\ \psi &\to \psi' = \exp[\tfrac{1}{4}ig(1+\gamma_5)(\tau \cdot \Lambda)]\psi = \{\tfrac{1}{2}(1+\gamma_5)\exp[\tfrac{1}{2}ig(\tau \cdot \Lambda)] + \tfrac{1}{2}(1-\gamma_5)\}\psi, \\ \sigma &\to \sigma' = \cos x\sigma + \frac{g}{2x}\sin x(\pi \cdot \Lambda), \quad (6.11) \\ \pi &\to \pi' = -\frac{g}{2x}\sin x\sigma\Lambda + \cos x\pi - \frac{g}{2x}\sin x(\Lambda \wedge \pi),\end{aligned}$$

where $x = \frac{1}{2}g\sqrt{(\Lambda \cdot \mathbf{A})}$. Notice here that the last three transformations all satisfy our condition. In terms of x the transformation for $\mathbf{u}_\mu$ can be written as follows:

$$\begin{aligned}S(g, \Lambda)\mathbf{u}_\mu &= \frac{g}{x}\sin x \cos x(\mathbf{u}_\mu \wedge \Lambda) + \frac{g^2}{2x^2}\sin^2 x\Lambda(\Lambda \cdot \mathbf{u}_\mu) + (1 - 2\sin^2 x)\mathbf{u}_\mu, \\ f(g, \Lambda)\partial_\mu \Lambda &= \frac{g}{2x^2}\sin^2 x(\partial_\mu \Lambda \wedge \Lambda) - \frac{g^2}{4x^2}\left(\frac{1}{x}\sin x \cos x - 1\right)\Lambda(\Lambda \cdot \partial_\mu \Lambda) \quad (6.12) \\ &\quad + \frac{1}{x}\sin x \cos x\partial_\mu \Lambda.\end{aligned}$$

Thus, eq. (6.10) is equivalent to the following Lagrangian (using eq. (6.6)):

$$\begin{aligned}\mathscr{L}_{G.L.} = \mathscr{L}'_{Y.M.} &- \bar{\psi}\{\gamma_\mu[\partial_\mu - \tfrac{1}{4}ig(1+\gamma_5)(\tau \cdot \mathbf{A}_\mu)] - h[\sigma + i(\tau \cdot \pi)\gamma_5]\}\psi \\ &- \tfrac{1}{2}(\partial_\mu \pi - \tfrac{1}{2}g\pi \wedge \mathbf{A}_\mu + \tfrac{1}{2}g\sigma \mathbf{A}_\mu)^2 - \tfrac{1}{2}(\partial_\mu \sigma - \tfrac{1}{2}g\pi \cdot \mathbf{A}_\mu)^2 \quad (6.10') \\ &- \tfrac{1}{2}\mu^2(\pi^2 + \sigma^2) - \lambda[\pi^2 + \sigma^2 - \tfrac{1}{4}f]^2 - \frac{\mu^2}{2f}\sigma',\end{aligned}$$

where $\sigma' = \cos x\,\sigma + (g/2x)\sin x\,(\pi \cdot \mathbf{b})$, with $x = \frac{1}{2}g\sqrt{\mathbf{b} \cdot \mathbf{b}}$.

Appendix

A.1. PROOF OF THEOREM I

Let $\phi_r = \phi_r(\phi'_s)$ be a point transformation. Hence we have

$$\phi_r = \sum_{n=0}^{\infty} c_n \phi'^n_s, \tag{A.1}$$

where c_n may contain $\phi'_t (t \neq s)$. Hence we get

$$\dot{\phi}_r = \sum_{n=0}^{\infty} \{c_n \underbrace{(\dot{\phi}'_s \phi'^{n-1}_s + \phi'_s \dot{\phi}'_s \phi'^{n-2}_s + \ldots + \phi'^{n-1}_s \dot{\phi}'_s)}_{n \text{ terms}} + \dot{c}_n \phi'^n_s\}. \tag{A.2}$$

Hence by definitions (3.9)—(3.12) one has

$$\begin{aligned} \frac{\partial}{\partial \dot{\phi}'_s} \dot{\phi}_r &= \sum_{n=0}^{\infty} c_n (\phi'^{n-1}_s + \phi'^{n-1}_s + \ldots + \phi'^{n-1}_s) = \sum_{n=0}^{\infty} c_n n \phi'^{n-1}_s \\ &= \frac{\partial}{\partial \phi'_s} \sum_{n=0}^{\infty} c_n \phi'^n_s = \frac{\partial}{\partial \phi'_s} \dot{\phi}_r \end{aligned} \tag{A.3}$$

as required.

A.2. PROOF OF THEOREM II

From eq. (A.2) and by using eqs. (3.9)—(3.12) it is easy to see that

$$\begin{aligned} \frac{\partial}{\partial \phi'_s} \dot{\phi}_r &= \sum_{n=0}^{\infty} [c_n \{\dot{\phi}'_s (n-1) \phi'^{n-2}_s + \phi'_s \dot{\phi}'_s (n-2) \phi'^{n-3}_s + \ldots + \phi'^{n-2}_s \dot{\phi}'_s \\ &\quad + \dot{\phi}'_s \phi'^{n-2}_s + \ldots + (n-2) \phi'^{n-3}_s \dot{\phi}'_s \phi'_s + (n-1) \phi'^{n-2}_s \dot{\phi}'_s\} + \dot{c}_n n \phi'^{n-1}_s] \\ &= \sum_{n=0}^{\infty} n c_n \{\dot{\phi}'_s \phi'^{n-2}_s + \phi'_s \dot{\phi}'_s \phi'^{n-3}_s + \ldots + \phi'^{n-3}_s \dot{\phi}'_s \phi'_s + \phi'^{n-2}_s \dot{\phi}'_s\} + \dot{c}_n n \phi'^{n-1}_s] \\ &= \sum_{n=0}^{\infty} \left[n c_n \frac{\partial}{\partial t} \{\phi'^{n-1}_s\} + \dot{c}_n n \phi'^{n-1}_s \right] \\ &= \frac{\partial}{\partial t} \sum_{u=0}^{\infty} c_n \frac{\partial}{\partial \phi'_s} \phi'^n_s = \frac{\partial}{\partial t} \frac{\partial}{\partial \phi'_s} \phi_r \end{aligned} \tag{A.4}$$

as required.

A.3. THE ENERGY-MOMENTUM TENSOR $T_{\mu\nu}$

Here we should like to prove $T_{\mu\nu} = T'_{\mu\nu}$ (eq. (3.16)). For a standard Lagrangian it is usual to define $T_{\mu\nu}$ as

$$T_{\mu\nu} = \sum_r \frac{1}{2} \left(\phi_{r,\mu} \frac{\partial \mathscr{L}}{\partial \phi_{r,\nu}} + \frac{\partial \mathscr{L}}{\partial \phi_{r,\mu}} \phi_{r,\mu} \right) - \delta_{\mu\nu} \mathscr{L}. \tag{A.5}$$

For a more general Lagrangian $\mathscr{L}$, however, this is inadequate, since, for

example, if $\mathscr{L}$ contains terms such as $f(\phi_r)\dot{\phi}_r^2$, the energy given by (A.5) is not conserved in general. To define $T_{\mu\nu}$ properly we first introduce a new operator [W $\partial/\partial Z$] which is defined analogously to $\partial/\partial Z$ (W, $Z = \phi_1$ or $\phi_2, \ldots$ or $\dot{\phi}_n$), by the following steps:

Step 1:

$$\left[W\frac{\partial}{\partial\phi_s}\right]\phi_r = \left[W\frac{\partial}{\partial\dot{\phi}_s}\right]\dot{\phi}_r = W\delta_{rs}, \qquad \left[W\frac{\partial}{\partial\dot{\phi}_r}\right]\phi_r = \left[W\frac{\partial}{\partial\phi_s}\right]\dot{\phi}_r = 0. \tag{A.6}$$

Step 2:

$$\left[W\frac{\partial}{\partial Z}\right](XY) = \left(\left[W\frac{\partial}{\partial Z}\right]X\right)Y + X\left(\left[W\frac{\partial}{\partial Z}\right]Y\right). \tag{A.7}$$

Note that this is not equal to $W\,\partial(XY)/\partial Z = W(\partial X/\partial Z)Y + WX\,\partial Y/\partial Z$, the difference being $[X, W]\partial Y/\partial Z$. The generalization of this is

$$\left[W\frac{\partial}{\partial Z}\right]\prod_{i=1}^{N}X_i = \sum_{j=1}^{N}\prod_{i=1}^{j-1}X_i\left(\left[W\frac{\partial}{\partial Z}\right]X_j\right)\prod_{i=j+1}^{N}X_i. \tag{A.8}$$

Step 3:

If F_i is a simple product as in Step 2 and $F = \sum_{i=1}^{M} F_i$, we put

$$\left[W\frac{\partial}{\partial Z}\right]F = \sum_{i=1}^{M}\left[W\frac{\partial}{\partial Z}\right]F_i. \tag{A.9}$$

The vital step is, of course Step 2. We can now define $T_{\mu\nu}$ in general. In fact,

$$T_{\mu\nu} \equiv \sum_r \left[\phi_{r,\mu}\frac{\partial}{\partial\phi_{r,\nu}}\right]\mathscr{L} - \delta_{\mu\nu}\mathscr{L}. \tag{A.10}$$

It is easily seen that this reduces to the usual definition when $\mathscr{L}$ contains the $\phi_{r,\mu}^2$ only with numerical coefficients.

Let us now prove the invariance of T_{44} under a point transformation (A.1). First we not that the standard $\mathscr{L}$ can be split into three parts,

$$\mathscr{L} = \mathscr{L}_{(2)} + \mathscr{L}_{(1)} + \mathscr{L}_{(0)}, \tag{A.11}$$

where $\mathscr{L}_{(2)}$, $\mathscr{L}_{(1)}$, $\mathscr{L}_{(0)}$ are bilinear, linear and of zero order, respectively, in the $\phi_{r,4}$. After making the point transformation we have, similarly,

$$\mathscr{L}'(=\mathscr{L}) = \mathscr{L}'_{(2)} + \mathscr{L}'_{(1)} + \mathscr{L}'_{(0)}, \tag{A.11'}$$

where $\mathscr{L}'_{(2)}$, $\mathscr{L}'_{(1)}$, $\mathscr{L}'_{(0)}$ are bilinear, linear, and of zero order, respectively, in the $\phi'_{r,4}$. The point nature of the transformation guarantees, however, that not only is $\mathscr{L} = \mathscr{L}'$, but also

$$\mathscr{L}_{(2)} = \mathscr{L}'_{(2)}, \qquad \mathscr{L}_{(1)} = \mathscr{L}'_{(1)}, \qquad \mathscr{L}'_{(0)} = \mathscr{L}'_{(0)}. \tag{A.12}$$

Next we note that by its definition (A.6)–(A.9), the operator $\sum_r[\phi_{r,4}\,\partial/\partial\phi_{r,4}]$ has the property that acting on a linear function of the $\phi_{r,4}$(e.g. $\phi^2\phi_{,4}+\phi_{,4}\phi^2$) it multiplies it by unity, and acting on a quadratic function (e.g $\phi\phi_{,4}{}^2+\phi_{,4}{}^2\phi$) it multiplies it by a factor 2. Hence we have

$$T_{44}=2\mathscr{L}_{(2)}+\mathscr{L}_{(1)}-\mathscr{L}, \qquad T'_{44}=2\mathscr{L}'_{(2)}+\mathscr{L}'_{(1)}-\mathscr{L}', \tag{A.13}$$

and therefore, by (A.12) one has

$$T_{44}=T'_{44}, \tag{A.14}$$

as required.

Similar proofs obtain for $T_{i4}=T'_{i4}$ and $T_{ij}=T'_{ij}$.

A.4. POINT TRANSFORMATIONS FOR FERMION VARIABLES

On account of the *anti*-commutation properties of fermion variables some new definitions of derivatives, etc. are needed in this case. We deal first with these new definitions.

Let X_r be a set of n anti-commuting quantities $\{X_s, X_r\}=0, r, s=1, 2, \ldots, n$. We consider variable transformations $X'_r(X_s)$ ot the X_r such that the X'_s can be expressed as sums of products of the X_r, with an *odd* number of X_r in each product, and conversely †. As a result one has

$$\{X_r, X_s\}=\{X_r, X'_s\}=\{X'_r, X'_s\}=0. \tag{A.15}$$

(Note that since $X_r{}^2=0$, each product can contain any given X_r at most linearly.) We next define drivatives of any functions F of the X_r, $\dot{X}_r$ and $X_r{}^\dagger$ which can be expressed as sums of products (not necessarily odd here) of these quantities ($X_r{}^\dagger$ means the Hermitian conjugate of X_r and is taken to be variationally independent of X_r) as follows:

Step 1:

$$\frac{\partial X_s}{\partial X_r}=\frac{\partial \dot{X}_s}{\partial \dot{X}_r}=\frac{\partial X_s{}^\dagger}{\partial X_r{}^\dagger}=\delta_{rs}, \qquad \frac{\partial X_r}{\partial X_s{}^\dagger}=\frac{\partial \dot{X}_r}{\partial X_s}=\frac{\partial \dot{X}_r{}^\dagger}{\partial \dot{X}_s}=0, \quad \text{etc.} \tag{A.16}$$

Step 2:

With $X, Y, Z, \ldots U=X_1, \ldots, X_r \ldots, \dot{X}_r, \ldots, X_r{}^\dagger, \ldots, \dot{X}_n{}^\dagger$ we put

$$\frac{\partial}{\partial U}(XYZ\ldots)=\frac{\partial X}{\partial U}YZ-X\frac{\partial Y}{\partial U}Z\ldots+\ldots \tag{A.17}$$

(note the minus sign: a plus sign would lead to the contradiction

$$2X_qX_p=\frac{\partial X_q}{\partial X_p}X_p{}^2+X_q\frac{\partial X_p}{\partial X_p}X_p+X_qX_p\frac{\partial X_p}{\partial X_p}=\frac{\partial}{\partial X_p}(X_qX_p{}^2)=\frac{\partial 0}{\partial X_p}=0\Big).$$

† An example of the sort of function we consider is $X'=X+XYZ$, $Y'=Y+YZX$, $Z'=Z+ZXY$; $X=X'-X'Y'Z'$, $Y=Y'-Y'Z'X'$, $Z=Z'-Z'X'Y'$

Step 3:

If the F_i are simple products as in Step 2, we put

$$\frac{\partial}{\partial X_s}\sum_i F_i = \sum_i \frac{\partial F_i}{\partial X_s}. \tag{A.18}$$

With these definitions we establish the following three theorems :

Theorem I':

$$\frac{\partial F(X'_s(X_r))}{\partial X_r} = \sum_s \frac{\partial X'_s}{\partial X_r}\frac{\partial F(X'_s)}{\partial X'_s}, \tag{A.19}$$

where X'_s is a sum of odd products of X_r.

Theorem II':

$$\frac{\partial}{\partial t}\frac{\partial}{\partial X_r}F(X_s) = \frac{\partial}{\partial X_r}\frac{\partial}{\partial t}F(X_s). \tag{A.20}$$

Theorem III':

$$\frac{\partial}{\partial \dot{X}_r}\frac{\partial}{\partial t}F(X_s) = \frac{\partial}{\partial X_r}F(X_s). \tag{A.21}$$

The last two are the analogues of theorems I and II of the boson case. The first is not as trivial as it looks and is not true unless the X'_s are sums of *odd* products of the X_r. All these theorems can be proved by induction and as the proof is rather simple, we shall not give it explicitly.

We are now in a position to consider the invariance of the Heisenberg-Pauli formalism under point transformations. Suppose we have a standard Lagrangian density

$$\mathscr{L} = -\tfrac{1}{2}\bar{\psi}(\gamma\overrightarrow{\partial}+\kappa)\psi+\tfrac{1}{2}\bar{\psi}(\gamma\overleftarrow{\partial}-\kappa)\psi+\bar{\psi}O\psi \tag{A.22}$$

with the usual notation, and where O is a function of the γ-matrices and some boson field only. The Heisenberg-Pauli equations $(\partial/\partial t)\,(\partial\mathscr{L}/\partial\dot{\phi}) = \delta\mathscr{L}/\delta\phi$ and $\{\partial\mathscr{L}/\partial\dot{\phi}\phi\} = i\Delta$ (Δ is a c-number) are in this case, according to our definitions (A.16)—(A.18), and regarding $\bar{\psi}$ and ψ as variationally independent, with $\partial\mathscr{L}/\partial\dot{\psi}_\alpha = -\frac{1}{2}i(\bar{\psi}\gamma_4)_\alpha$, $\partial\mathscr{L}/\partial\dot{\bar{\psi}}_\alpha = -\frac{1}{2}i(\gamma_4\psi)_\alpha$,

$$\begin{aligned}
&\frac{\partial}{\partial t}\frac{1}{2i}(\bar{\psi}\gamma_4)_\alpha = \kappa\psi_\alpha-(\bar{\psi}\gamma_i\overleftarrow{\partial}_i)_\alpha-(\bar{\psi}0)_\alpha-\frac{1}{2i}(\dot{\bar{\psi}}\gamma_4)_\alpha, \qquad \left\{\frac{1}{2i}(\bar{\psi}\gamma_4)_\alpha, \psi_\beta\right\} = i\Delta\delta_{\alpha\beta},\\
&\frac{\partial}{\partial t}\frac{1}{2i}(\gamma_4\psi)_\alpha = -\kappa\psi_\alpha-(\gamma_i\overrightarrow{\partial}_i\psi)_\alpha+(0\psi)_\alpha-\frac{1}{2i}(\gamma_4\dot{\psi})_\alpha, \quad \left\{\frac{1}{2i}(\gamma_4\psi)_\alpha, \bar{\psi}_\beta\right\} = i\Delta\delta_{\alpha\beta},
\end{aligned} \tag{A.23}$$

which are the usual equations for a spinor field. We now make a point transformation, which we define to be

$$\psi_\alpha = \psi'_\alpha F(\psi'_\beta), \tag{A.24}$$

where F is a function only of bilinear forms of ψ'. [F does not contain $\dot{\psi}'$,

just as in the boson case, but in addition F does not contain $\bar{\psi}'$, which is a restriction peculiar to the fermion case and is imposed because $\{\psi, \bar{\psi}\} \neq 0$. Covariant bilinear forms of ψ' alone can be constructed by means of the B-matrix of Pauli [13]); for example: $\bar{\psi}B\gamma_\mu\chi$ is a vector.] The transformation (A.24) is automatically of the type stipulated above (assuming, of course, that F can be expressed as a power series in its argument) and theorems I', II' and III' apply.

We now prove the invariance of the Heisenberg-Pauli formalism (A.23) under the transformation (A.24). From (A.22) we get, using theorem III',

$$\frac{\partial \mathscr{L}}{\partial \dot{\psi}'_\alpha} = \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial \dot{\psi}_\beta}{\partial \dot{\psi}'_\alpha} + \frac{1}{2i}\frac{\partial \dot{\bar{\psi}}_\beta}{\partial \dot{\psi}'_\alpha}(\gamma_\alpha \psi)_\beta = \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial \dot{\psi}_\beta}{\partial \dot{\psi}'_\alpha} = \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial \psi_\beta}{\partial \psi'_\alpha}.$$

Thus, one has, according to eq. (A.23), theorem II' and a slight extension of theorem I' to cover functional derivatives,

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial \mathscr{L}}{d\dot{\psi}'_\alpha} &= \left[\frac{\partial}{\partial t}\frac{1}{2i}(\bar{\psi}\gamma_4)_\beta\right]\frac{\partial \psi_\beta}{\partial \psi'_\alpha} + \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial}{\partial t}\left(\frac{\partial \psi_\beta}{\partial \psi'_\alpha}\right) \\ &= \frac{\delta \mathscr{L}}{\delta \psi_\beta}\frac{\partial \psi_\beta}{\partial \psi'_\alpha} + \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial}{\partial t}\left(\frac{\partial \psi_\beta}{\partial \psi'_\alpha}\right) = \frac{\delta \mathscr{L}}{\delta \psi_\beta}\frac{\partial \psi_\beta}{\partial \psi'_\alpha} + \frac{1}{2i}(\bar{\psi}\gamma_4)_\beta \frac{\partial \dot{\psi}_\beta}{\partial \dot{\psi}'_\alpha} \\ &= \frac{\delta \mathscr{L}}{\delta \psi_\beta}\frac{\partial \psi_\beta}{\partial \psi'_\alpha} + \frac{\partial}{\partial \psi'_\alpha}\left(\frac{-1}{2i}\bar{\psi}\gamma_4\dot{\psi}\right) = \frac{\delta \mathscr{L}}{\delta \psi'_\alpha}.\end{aligned}$$

On the other hand, since $[\partial \psi_\gamma / \partial \dot{\psi}'_\alpha, \psi'_\beta] = 0$, and using theorem I', one has

$$\begin{aligned}\left\{\frac{\partial \mathscr{L}}{\partial \dot{\psi}'_\alpha}, \psi'_\beta\right\} &= \left\{\frac{1}{2i}(\bar{\psi}\gamma_4)_\gamma \frac{\partial \psi_\gamma}{\partial \dot{\psi}'_\alpha}, \psi'_\beta\right\} = \frac{1}{2i}\{(\bar{\psi}\gamma_4)_\gamma, \psi'_\beta\}\frac{\partial \psi_\gamma}{\partial \dot{\psi}'_\alpha} \\ &= \frac{\partial \psi'_\beta}{\partial \psi_\gamma}\frac{\partial \psi_\gamma}{\partial \dot{\psi}'_\alpha} i\Delta = \frac{\partial \psi'_\beta}{\partial \dot{\psi}'_\alpha} i\Delta = \delta_{\beta\alpha} i\Delta.\end{aligned}$$

Similar equations obtain for $\partial \mathscr{L} / \partial \dot{\bar{\psi}}'_\alpha$. The invariance of $T_{\mu\nu}$ can be established here in the same way as in the boson case. Thus, the invariance of the Heisenberg-Pauli formalism is established.

Here, we have considered only one fermion field, but it is very probable that the above considerations can be extended to the case of several fermion fields, possibly interacting directly (as in the case of universal Fermi interactions).

The last question we must consider is a point transformation which mixes fermion and boson variables. For simplicity, and because it is only cases of this kind which we have had to consider in applications, we take a Lagrangian of the type

$$\mathscr{L} = -\tfrac{1}{2}\bar{\psi}(\gamma\overrightarrow{\partial} + \kappa)\psi + \tfrac{1}{2}\bar{\psi}(\gamma\overleftarrow{\partial} - \kappa)\psi + g\bar{\psi}0\psi + \tfrac{1}{2}(\dot{\phi}^2 + (\nabla\phi)^2 + \mu^2\phi^2). \quad \text{(A.25)}$$

It is then easy to establish along the lines of section 3 and of this Appendix that the Heisenberg-Pauli formalism for this Lagrangian is invariant under the point transformation $\psi_\alpha = \psi'_\alpha F(\psi', \phi')$, $\phi = f(\phi')$, where F is a function of ϕ' and of bilinear forms of ψ' and F and f are, as usual, expressible as (possibly infinite) power series of their arguments.

The authors would like to thank Professor Y. Takahashi for helpful discussions. One of them (S. K.) is indebted to Professor J. L. Synge for kind hospitality at his institute.

References

1) J. S. R. Chisholm, Nuclear Physics, **26** (1961) 469
2) R. Haag, Nuovo Cim. Suppl. **14** (1959) 131;
M. Dresden, Lectures in theoretical physics, Vol. II (Interscience Publ., New York, 1959) 366
3) H. Umezawa and Y. Takahashi, Prog. Theor. Phys. **9** (1953) 14,501;
Y. Katayama, Prog. Theor. Phys. **10** (1953) 31;
W. Pauli, Nuovo Cim. **10** (1953) 648
4) W. Heisenberg and W. Pauli, Zeit. f. Phys. **56** (1929) 1, **59** (1929) 168
5) Z. Koba, T. Tati and S. Tomonaga, Prog. Theor. Phys. **3** (1948) 141, 229;
J. Schwinger, Phys. Rev. **74** (1948) 1439
6) E. C. Nelson, Phys. Rev. **60** (1941) 830;
F. J. Dyson, Phys. Rev. **73** (1948) 929;
K. M. Case, Phys. Rev. **76** (1949) 14
7) A. Salam, Nuclear Physics **18** (1960) 681;
S. Kamefuchi, Nuclear Physics **18** (1960) 691
8) H. Umezawa and S. Kamefuchi, Nuclear Physics **23** (1961) 399
9) S. R. Drell and E. F. Henley, Phys. Rev. **88** (1952) 1053
10) C. N. Yang and R. L. Mills, Phys. Rev. **96** (1954) 191
11) F. Gürsey, private communication
12) M. Gell-Mann and M. Lévy, Nuovo Cim. **16** (1960) 705;
A. Salam and J. C. Ward, Nuovo Cim. **19** (1961) 165;
A. Peterman and H. Ruegg, Helv. Phys. Acta **33** (1960) 143
13) W. Pauli, Inst. H. Poincare Ann. **6** (1936) 109

Reprinted from THE PHYSICAL REVIEW, Vol. 130, No. 3, 1287-1290, 1 May 1963

Renormalizable Electrodynamics of Vector Mesons

ABDUS SALAM
Imperial College, London, England
(Received 26 December 1962)

It is shown that the conventional theory of charged spin-one mesons interacting with photons can be renormalized provided meson mass m and charge e are restricted by the relation $Z(m^2,e^2)=0$, where Z is the meson wave-function renormalization constant.

1. INTRODUCTION

IT is the purpose of this paper to show that conventional theory of charged spin-one mesons interacting with photons (vector electrodynamics) can be renormalized provided the meson mass m and the physical coupling constant e are restricted by the relation

$$Z(m^2,e^2)=0, \tag{1}$$

where Z is the meson wave-function renormalization constant. The essence of the proof lies in showing that if (1) is satisfied, the modified vertex function Γ_1 behaves like $\sim 1/p$ and the modified propagator Δ_1 like ~ 1 for large p. Thus if S-matrix elements are computed as in Dyson's method, by first drawing irreducible diagrams and then writing Δ_1 and Γ_1 for each line and each vertex, the resulting integrals are all finite (except possibly those for meson self-mass and photon wave-function renormalization constant).

In Secs. 2 and 3 the necessary formalism is developed; Sec. 4 outlines the proof and Sec. 5 is concerned with the implications of relation (1). In a separate paper with R. Delbourgo we give actual computations of Δ_1 and Γ_1.

2. THE PROPAGATOR

Let the (renormalized) fields $A_\mu^\pm$ describe charged stable vector particles of mass m. The conventional Lagrangian for vector electrodynamics is

$$L=-\tfrac{1}{2}ZF_{\mu\nu}{}^+F_{\mu\nu}{}^- - Zm_0{}^2A_\mu{}^+A_\mu{}^- - \tfrac{1}{4}F_{\mu\nu}{}^0F_{\mu\nu}{}^0, \tag{2}$$

where

$$F_{\mu\nu}{}^\pm=\partial_\mu{}^\pm A_\nu{}^\pm-\partial_\nu{}^\pm A_\mu{}^\pm,$$
$$\partial_\mu{}^\pm=\partial/\partial x_\mu \mp ieA_\mu{}^0.$$

$A_\mu{}^0$ is the photon field and Z and $m_0{}^2$ are constants specified below.

We write the Fourier transform of the propagator[1] $\langle A_\mu{}^+(x)A_\nu{}^-(y)\rangle_+$ in the form

$$\Delta_{1\mu\nu}(p)=d_{\mu\nu}\lambda_1(p^2)+e_{\mu\nu}\lambda_2(p^2), \tag{3}$$

where[2]

$$d_{\mu\nu}=(-\delta_{\mu\nu}+p_\mu p_\nu/p^2), \tag{4}$$

$$e_{\mu\nu}=p_\mu p_\nu/p^2. \tag{5}$$

By hypothesis the spin-one part of Δ_1 has a pole at $p^2=m^2$ with "residue" $d_{\mu\nu}$. Thus $\lambda_1(p^2)$ must have the form

$$\lambda_1{}^{-1}(p^2)=(p^2-m^2)Z(p^2), \tag{6}$$

where $Z(p^2)$ equals

$$Z(p^2)=1-(p^2-m^2)\int \frac{G_1(K^2)dK^2}{p^2-K^2+i\epsilon}. \tag{7}$$

Also the condition that there is no pole at $p^2=0$, means

$$\lambda_1(0)+\lambda_2(0)=0. \tag{8}$$

One may, therefore, write[3]:

$$\lambda_2{}^{-1}(p^2)=-\lambda_1{}^{-1}(0)-m^2p^2\int \frac{G_2(K^2)dK^2}{p^2-K^2+i\epsilon}. \tag{9}$$

Note that with (6) and (9)

$$\lim_{p^2\to m^2}(p^2-m^2)\Delta_{1\mu\nu}=(-\delta_{\mu\nu}+p_\mu p_\nu/m^2).$$

We now define the constants Z and $m_0{}^2$ which occur in the Lagrangian. Let[4]

$$Z=\lim_{p^2\to\infty}Z(p^2)=1-\int G_1dK^2, \tag{10}$$

$$Zm_0{}^2=\lim_{p^2\to\infty}\lambda_2{}^{-1}(p^2)=m^2\left(1-\int\frac{m^2}{K^2}G_1-\int G_2\right). \tag{11}$$

[1] We follow the notation of the excellent paper by K. W. Ford, Nuovo Cimento 24, 1671 (1962).

[2] Writing d and e for $d_{\mu\nu}$ and $e_{\mu\nu}$, note that $-e+d=-1$, $dd=-d$, $ee=e$, $de=ed=0$. Also if $\Delta=\lambda_1 d+\lambda_2 e$, then $\Delta^{-1}=\lambda_1{}^{-1}d+\lambda_2{}^{-1}e$.

[3] For theories where conservation laws of the type $\partial J^\pm/\partial x_\mu=0$ hold, $G_2=0$ and λ_2 is a constant. This clearly is not the case for the present theory.

[4] Canonical commutation relations give alternative (but equivalent) expressions for Z and $Zm_0{}^2$. Thus comparing the canonical values of $[A_k(\mathbf{x}),A_l(\mathbf{x}')]$ and $[A_k(\mathbf{x}),\dot{A}_l{}^-(\mathbf{x}')]$, $(k,l=1,2,3)$ with those deduced from dis. $\Delta_1(x-x')$ one obtains:

$$Z^{-1}=1+\int G_1(K^2)\,|Z(K^2)|^{-2}dK^2, \tag{A}$$

$$\frac{1}{Zm_0{}^2}=\frac{1}{m^2}\int[G_1(K^2)\,|Z(K^2)|^{-2}+K^2G_2(K^2)\,|\lambda_2(K^2)|^2m^2]\frac{dK^2}{K^2}. \tag{B}$$

To establish equivalence of (10) with (A) for example, note that

$$\mathrm{Im}Z^{-1}(p)=|Z(p^2)|^{-2}G(p^2),$$

so that,

$$Z^{-1}(p^2)=1+(p^2-m^2)\int\frac{G_1|Z(K^2)|^2}{p^2-K^2+i\epsilon}dK^2,$$

provided the last integral converges. Comparison with (7) in the limit $p^2\to\infty$ proves the equivalence of (10) and (A).

The canonical expression for $[A_k{}^+(\mathbf{x}),\dot{A}_l{}^-(\mathbf{x}')]$, has been implicitly given by G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., New York), p. 93. To obtain it use the identity

$$\dot{A}_l{}^\pm=Z^{-1}[\pi_l{}^m-(1/m_0{}^2)\partial_l{}^\pm\partial_j{}^\pm\pi_j{}^m]-ieA_4{}^0A_l{}^\pm, \tag{C}$$

which can be derived from the equations of motion. The canonical momenta π_j occurring in (C) satisfy,

$$[A_k{}^\pm(\mathbf{x}),\pi_j{}^\pm(\mathbf{x}')]=i\delta_{ij}\delta(\mathbf{x}-\mathbf{x}').$$

Thus finally

$$\lambda_2^{-1}(p^2)=Zm_0^2-m^2\int\frac{K^2G_2(K^2)}{p^2-K^2+i\epsilon}. \quad (12)$$

3. THE VERTEX FUNCTION

For the vertex function $\Gamma_{\mu\nu 1}{}^{\alpha}(p,p')$, by considering the product $\Box_z(\partial/\partial z_\alpha)\langle A_\mu{}^+(x)A_\nu{}^-(y)A_\alpha{}^0(z)\rangle_+$, one deduces the Ward-Takahashi identity

$$\Delta_1^{-1}(p)-\Delta_1^{-1}(p')=-(p-p')_\alpha\Gamma_1{}^\alpha(p,p') \quad (13)$$

and its differential form

$$\partial\Delta_1^{-1}/\partial p_\alpha=-\Gamma_1{}^\alpha(p,p). \quad (14)$$

Also from charge-conjugation invariance

$$\Gamma_{\mu\nu 1}(p,p')=-\Gamma_{\nu\mu 1}{}^\alpha(-p',-p). \quad (15)$$

Equation (13) can be solved to give

$$\Gamma_1=\Gamma_A+\Gamma_B.$$

Here Γ_B is an arbitrary function which satisfies[6] $(p-p')_\alpha\Gamma_B{}^\alpha=0$, and

$$-\Gamma_A=\frac{(p+p')_\alpha}{p^2-p'^2}[\Delta^{-1}(p)-\Delta^{-1}(p')]. \quad (16)$$

Explicitly,

$$\Delta^{-1}(p)-\Delta^{-1}(p')=(p^2-p'^2)A_1(p^2p'^2)+[p_\mu p_\nu X(p^2)-p_\mu'p_\nu'X(p'^2)], \quad (17)$$

where

$$A_1(p^2,p'^2)=Z+\int\frac{(K^2-m^2)^2G_1}{(p^2-K^2)(p'^2-K^2)}dK^2,$$
$$X(p^2)=Z-\int\frac{(K^2-m^2)^2G_1}{K^2(p^2-K^2)}dK^2-m^2\int\frac{G_2dK^2}{p^2-K^2}. \quad (18)$$

In general, all integrals involved in Δ and Γ_A converge provided[6]

$$\int G_1dK^2<\infty,\quad \int G_2dK^2<\infty. \quad (19)$$

Now if e and m are so related that[7]

$$Z(e^2,m^2)=0,$$

and if we make (at this stage, *ad hoc*) assumption that

$$\lim_{K^2\to\infty}G_1\sim(1/K^2)^2, \quad (20)$$

Eq. (18) shows that for large p or p', A_1, X, etc., have the form

$$A_1(p^2,p'^2)\sim 1/(ap^2+bp'^2),$$
$$X(p^2)\sim 1/p^2,$$

and therefore

$$\Gamma_1\sim 1/(\alpha p+\alpha' p'). \quad (21)$$

The same conditions ensure that

$$Z(p^2)\sim 1/p^2 \quad \text{and} \quad \Delta_1\sim 1. \quad (22)$$

In the Sec. 4, we consider the validity of (20).[8]

4. INTEGRAL EQUATIONS FOR Γ_1 AND Δ_1

To set up the coupled equations Δ_1 and Γ_1 and to write general scattering matrix elements, we follow Dyson's method and split off from L the conventional free Lagrangian L_0 which forms the basis of the interaction representation.[9] {Notice that the interaction Lagrangian contains (nondivergent) self-mass terms as well as kinetic energy terms of the type $(Z-1)A_\mu{}^+\times[(p^2-m^2)d_{\mu\nu}+m^2e_{\mu\nu}]A_\nu{}^-$.} Instead of writing S-matrix elements in terms of the free propagator

$$\Delta_{F0}=\mathbf{d}/(p^2-m^2)+(1/m^2)\mathbf{e}$$

and the unmodified vertex,[10]

$$\Gamma_0=\delta_{\mu\nu}(p+p')_\alpha-\delta_{\mu\alpha}p_\nu-\delta_{\nu\alpha}p_\mu',$$

we first compute Δ_1 and Γ_1 as solutions of the integral equations below which are derived from the given Lagrangian and then write down other S-matrix elements by drawing irreducible graphs and by inserting in these Δ_1 and Γ_1 for the lines and the vertices.

The integral equations for Γ_1 and Δ_1 are

$$\Gamma_1(p,p')=Z\Gamma_0(p,p')+K(p,p'), \quad (23)$$

where

$$K=e^2\int\Gamma_1(e)\Delta_1(e)\Gamma_1(e)\Delta_1(e)\Gamma_1(e)D_1(e)+e^4\int\cdots, \quad (24)$$

---

[6] Γ_B [which contains the dependence of Γ_1 on $(p-p')=t$] must have the general form:

$$\Gamma_B=[t_\alpha(p^2-p'^2)-t^2(p+p')_\alpha]\times[\delta_{\mu\nu}F_1+p_\mu p_\nu'F_2+p_\mu t_\nu F_3+p_\nu't_\mu F_3'+t_\mu t_\nu F_4]+(\delta_{\alpha\mu}t_\nu-\delta_{\alpha\nu}t_\mu)F_5+(t^2\delta_{\alpha\mu}-t_\alpha t_\mu)p_\nu'F_6-(t^2\delta_{\alpha\nu}-t_\alpha t_\nu)p_\mu F_6'+(t^2\delta_{\alpha\mu}-t_\alpha t_\mu)t_\nu F_7-(t^2\delta_{\alpha\nu}-t_\alpha t_\nu)t_\mu F_7',$$

where, using (15), the invariant functions F_1, F_2, F_4, F_5 are symmetric in p and p', and for F_3, F_6, and F_7, $F'(p^2,p'^2,t^2)=F(p'^2,p^2,t^2)$. F_5 gives the magnetic moment and F_4 the quadrupole moment of the vector particle. Further on we make the approximation $\Gamma\approx\Gamma_A$.

[6] This means both Z and Zm_0^2 are finite. Note that m_0^2 always occurs multiplied by the constant Z.

[7] With zero-photon mass there is no mass other than m in the present theory. Thus, the relation must reduce to $s(e^2)=0$.

[8] All these statements are accurate to the extent that powers of $(\ln p^2)$ are ignored.

[9] For details of the procedure see P. T. Matthews and A. Salam, Phys. Rev. 94, 185 (1954). One would get the same Eqs. (23) and (25) if Schwinger's Green's function method is used with the Z-containing Lagrangian (2), [J. Schwinger, Proc. Natl. Acad. Sci. 37, 452 (1951)].

[10] Throughout this paper we have consistently ignored the so-called "Compton parts," i.e., the modifications of the 2-meson 2-photon vertices which occur in electrodynamics of bose particles. Since Ward-Takahashi identities hold also for these graphs, their high-energy behavior presents no new conceptual difficulties. In this respect the β formalism for vector electrodynamics would have been superior to the formalism of this paper because no "Compton part" insertions are necessary in that case.

and[11,12]

$$\Delta_{\mu\nu 1}^{-1}(p) = -(p-p')_\alpha \times \int_0^1 \Gamma_{\mu\nu 1}{}^\alpha(px+p'(1-x), px+p'(1-x))dx \quad (25)$$

with $p'^2=m^2$, and all terms involving $p_\mu{}^1$, p_ν' omitted.

To solve (23) and (25), first consider the case $Z\neq 0$. The following approximation procedure reproduces the conventional perturbation series: (i) Take $Z\Gamma_0(p,p')$ as the first approximation to Γ_1. (ii) Integrate Eq. (25); the first approximation to Δ_1^{-1} therefore is $Z\Delta_{F0}^{-1}$. Since $\Delta_1 \mathrm{d}$ has a pole at $p^2=m^2$ with residue $=\mathrm{d}$, to this order $Z=1$. (iii) Use Γ_0 and Δ_{F0} in (24) to obtain the next approximation[13] to Γ_1, Δ_1 (and Z), and so on.

Since $\Gamma_0\sim(p+p')$, it is clear that for the inhomogeneous case ($Z\neq 0$), Γ_1 is unlikely to converge faster than $(p+p')$.

If $Z=0$, we show below that the situation so far as the high-energy behavior is concerned is completely different. However, one may still set up an approximation scheme similar to the above, with only the change that in the zeroth approximation $Z\Delta_{F0}^{-1}$ is to be replaced by a suitable $(\Delta_1^{(0)})^{-1}$. Thus, in an obvious notation: (i) Take $\Gamma_1^{(0)}=\Gamma_A$ as defined in (17) with two unknown functions λ_1 and λ_2. Integrate (25) to get $\Delta_1{}^{(0)}=\lambda_1\mathrm{d}+\lambda_2\mathrm{e}$. (ii) Insert Γ_A for Γ_1 on the right hand side of (24) (fixing for practical purposes on some suitable subset of irreducible graphs). This gives:

$$\Gamma_1(p,p')=\Gamma_A(p,p')+\Gamma_B(p,p')=K[\Gamma_A]. \quad (26)$$

For $p=p'$, $\Gamma_B=0$. Thus,

$$\Gamma_A(p,p)=\frac{\partial}{\partial p}\Delta^{-1}(p)=K[\Gamma_A]_{p=p'}. \quad (27)$$

This is a set of homogeneous equations for λ_1 and λ_2 or, equivalently, G_1 and G_2. Once these are solved, (26) for $p\neq p'$ gives the zeroth approximation to Γ_B; the next approximations to G_1, G_2, and Γ_B are obtained by successive substitutions in $K[\Gamma_1]$.

The crucial step for the entire procedure then is the initial determination of $G_1{}^{(0)}$ and $G_2{}^{(0)}$ solutions of (27).

In a second paper (with R. Delbourgo) we present these solutions and show that $G_1{}^{(0)}$ and $G_2{}^{(0)}$ do indeed satisfy the convergence criteria of (19) and (20).[14] Here we show that if the initial approximations $G_1{}^{(0)}$, $G_2{}^{(0)}$ satisfy (19) and (20), all successive approximations possess the same property, and that $\Gamma_1(p,p')$ falls essentially as $1/p$ (or $1/p'$) when either of the variables p or p' is large. The proof is elementary. Since $\Delta_1{}^{(0)}(p)\sim 1$ by hypothesis, we infer from (17) that $\Gamma_A{}^{(0)}\sim 1/p$. Assuming that the photon propagator[15] $D_1(p^2)\sim 1/p^2$, one can see that the integral on the right of (26) must converge, yielding $\Gamma_1{}^{(1)}\approx 1/(\alpha p+p'\alpha')$ so far as dimensions are concerned. From (25) this means that the next approximation is $\Delta_1{}^{(1)}(p)\sim 1$,[15] and that $G_1{}^{(1)}$ and $G_2{}^{(1)}$ satisfy (19) and (20).

Before closing this section, we estimate the dimensional behavior of any Feynman integral with E_m external meson and E_p external photon lines. With $\Gamma\sim 1/p$ and $\Delta_1\sim 1$, these integrals converge provided[16,17]

$$2E_m+E_p>4.$$

---

[11] Equation (25), which is the integral equivalent of (13), was first derived by J. C. Ward, Phys. Rev. 84, 897 (1951). Instead one may work with the Dyson equation:

$$\Delta^{-1}=[Z(p^2-m_0{}^2)\mathrm{d}+Zm_0{}^2\mathrm{e}]-\pi_1{}^*(p), \quad (A)$$

where

$$\pi_1{}^*=\sum_{\substack{\text{sum over categories}\\ \text{of graphs}}} e^2\int \Gamma_0\Delta_1 D_1\Gamma_0+\cdots. \quad (B)$$

"Categories" which here take the place of "irreducible" graphs have been defined by A. Salam, Phys. Rev. 82, 217 (1951). Equation (A) is more general in so far as it applies also to nongauge-invariant theories. However, its disadvantage is the explicit appearance of Γ_0 on the right-hand side of (B). As is well known from the analysis of the "overlapping" self-energy parts $\pi_1{}^*$ has the same behavior as $\sim\int\Gamma_1\Delta_1\Gamma_1 D_1$.

[12] More precisely, one should also write an integral equation for the photon propagator $D_1(p)$ and solve it simultaneously with the equations for Γ_1 and Δ_1. For purposes of the present paper this is an unnecessary complication.

[13] To maintain gauge invariance and for consistency with (25), the set of irreducible graphs retained at each stage of approximation should include also graphs made up from appropriate "Compton-parts." These points will be covered in a second paper.

[14] The chief difficulty of solving (27) lies in the imposition of $Z(e^2,m^2)=0$. This is because Z itself is being computed (in terms of G_1) at the same time as the equation is being solved. It is worth noticing that it is only the real part of Δ^{-1} or Γ [see Eq. (23) or Eq. (A) of footnote 11] which explicitly depends on Z. Thus it is the high-energy behavior of only the real part which is in error unless we use $Z(e^2,m^2)=0$.

[15] This also means $Z^{(1)}(p^2)\sim 1/p^2$ so that

$$Z^{(1)}=\lim p^2\to\infty Z(p^2)=0.$$

In other words, one does not improve on e^2 deduced from $Z^{(0)}(e^2,m^2)=0$ unless more irreducible graphs are included in the approximation to K.

[16] To see how this works out in practice, consider the simple case of a single closed loop with n external photon lines ($n>4$) of momenta $k_1,\cdots,k_n$. The Feynman integral has the form:

$$F(k_1,\cdots,k_n)=\int d^4p\,[\Gamma_1(p,p+k)]^n[\Delta_1(p+k)]^n.$$

For large p, the behavior of the integrand is dominated by the basic unit

$$\Gamma_1(p,p)\Delta_1(p)=\left(\frac{\partial}{\partial p}\Delta_1{}^{-1}(p)\right)\Delta_1(p) = (\lambda_1'\mathrm{d}+\lambda_2'\mathrm{e}+\lambda_1\mathrm{d}'+\lambda_2\mathrm{e}')(\lambda_1{}^{-1}\mathrm{d}+\lambda_2{}^{-1}\mathrm{e}) \approx (1/p)+(1/p)(\lambda_2/\lambda_1+\lambda_1/\lambda_2).$$

For unmodified Δ_{F0} and Γ_0, $\lambda_1\sim 1/p^2$, $\lambda_2\sim 1$ so that the basic unit $\Gamma_0\Delta_{F0}$ has the behavior

$$\Gamma_0\Delta_{F0}\approx p.$$

If, however, (19)–(20) are satisfied, so that λ_1, $\lambda_2\sim 1$ then

$$\Gamma_1\Delta_1\sim 1/p$$

and the closed-loop integral $F(k_1,\cdots,k_n)$ ($n>4$) is convergent. The moral of the above analysis for the renormalization of electrodynamics of higher spin particles is clear. For a Yukawa-type theory, renormalizability needs an increase rate of $\Gamma_1\Delta_1$ no faster than $1/p$. A necessary condition for this seems to be that neither one of the two "orthogonal" functions λ_1 and λ_2 (of which Δ_1 is made up) should be more convergent than the other.

[17] For the conventional theory of scalar mesons interacting with photons and with no relation like $Z(e^2,m^2)=0$ operative, the

This means that the only possible infinite integrals are those corresponding to photon wave-function renormalization and meson self-mass[18] and all other S-matrix elements are finite.

5. THE RELATION $Z(e^2,m^2)=0$

It is the contention of this paper that a sensible vector electrodynamics exists for some special values of meson mass and charge.[19]

In considering the implications of $Z=0$ it is perhaps instructive to clarify the relationship of the field $A(x)$ to the so-called unrenormalized field $A_u(x)=Z^{+1/2}A(x)$. One could perfectly well rewrite the entire theory formally in terms of $A_u(x)$ so that the Z factors disappear from the Lagrangian. One may now require $Z(e^2)=0$ but the important remark is that even if this is the case, unlike (23) and (25), the equations satisfied by Δ_u^{-1} and Γ_u are not homogeneous. Thus two distinct situations may[20] be envisaged:

(A) Field $A_u(x)$ describes the physical situation. The propagator Δ_u has no pole ($Z=0$). Thus there is no stable physical particle and the conventional measurement of meson's electric charge e (using limiting static electric fields) presents conceptual difficulties. Most important of all, there is no dimunition in the divergence of the theory.

(B): Field $A(x)$ describes a stable particle of mass m. In this case the unrenormalized fields and particles have no meaning whatever. The "true" field $A(x)$ may correspond possibly to nonelementary spin-one particles (like the deuteron) and a restriction on mass and charge $Z(e^2,m^2)=0$ is a necessity for the theory to make sense. The role of conditions like $Z=0$ in connection with theories of composite particles has been discussed earlier.[21] It was, of course, not appreciated then, that the same condition would also prove necessary for renormalizability[22] for spins other than 0 and $\frac{1}{2}$.

For interaction of spin-one particles with fermions, it is clear that besides the vanishing of the meson wave-function renormalization constant, ($Z_3=0$) we shall also need $Z_1=0$ where Z_1 renormalizes the vertex part. Thus there must be a functional relationship between the coupling constant and fermion and meson masses.

In reference (20) we envisaged theories with $Z_1=Z_2=Z_3=0$. In a future paper the convergence properties of integrals in such theories are investigated. We conjecture that such field theories have no infinities whatever and that quantum theory of fields is a subject wrongfully, unduly, and much maligned in the past, principally by its friends.

The author is indebted to S. Kamefuchi, G. Feldman, P. T. Matthews, and L. Brown for stimulating discussions.

corresponding condition is $E_u+E_p>4$. Unlike the case of scalar electrodynamics, meson-meson scattering seems to be convergent for spin-one particles.

[18] Since m_u^2 always occurs in the combination Zm_u^2, for $Z=0$, a finite m_u^2 would imply a second relation between e^2 and m^2, i.e.,

$$\int \frac{K^2-m^2}{K^2}G_1(K^2)dK^2=\int G_2(K^2)dK^2,$$

which is unlikely to be true in general. [It certainly does not hold for an electrodynamics of the type suggested by A. Salam and J. C. Ward [Nuovo Cimento 11, 568 (1959)] where $\partial j_\mu^{\pm}/\partial x_\mu=0$ and, therefore, $G_2=0$.]

[19] Thus an expansion around $e=0$, is unthinkable. It is perhaps in this sense that there may be some correspondence between the present paper and recent work of T. D. Lee on renormalization of vector-electrodynamics where matrix elements are shown to depend on $e\ln e$. [T. D. Lee, Phys. Rev. 128, 899 (1962); see also C. N. Yang and T. D. Lee, 128, 855 (1962)].

[20] I am indebted to Professor G. Feldman for the following simple antithesis between Cases (A) and (B). Since $Z\approx|\langle\text{bare}|\text{true}\rangle|^2$, $Z=0$ means: either (A) $|\text{true}\rangle=0$, $|\text{bare}\rangle\neq0$; no true stable particle exists and there is no "renormalized" field; or (B) $|\text{bare}\rangle=0$; $|\text{true}\rangle\neq0$; there is no "elementary" field $A_u(x)$; however the "true" field $A(x)$ exists and corresponds possibly to a composite particle.

[21] A. Salam, Nuovo Cimento 25, 224, 1962; 1962 International Conference on High Energy Physics, CERN, p. 686. S. Weinberg *Op. cit.* p. 683. M. J. Vaughan, R. Aaron, and R. D. Amado, Phys. Rev. 124, 1258 (1961). J. C. Howard, and B. Jouvet, Nuovo Cimento 18, 466, 1960. R. Acharya, Rochester preprint NYO 10125. S. Weinberg, Phys. Rev. 130, 776 (1963). I am indebt to Professor Steven Weinberg for stressing to me the virtues of "zero Lagrangians."

[22] In an early paper S. F. Edwards, Phys. Rev. 90, 282 (1953) did point out that $Z=0$ is a necessary condition for the solution of $\Gamma_1=Z+K[\Gamma_1]$ for the electrodynamics of spin-$\frac{1}{2}$ particles. In his paper, however, the coupled equation for Δ^{-1} was not simultaneously considered so that it is somewhat tricky to compare his results with ours. When dealing with non-gauge-invariant theories where no Ward identity exists, it is, of course, unlikely that for the basic unit $\Gamma_1\Delta_1\sim1/p^\alpha$, α will equal 1, and a more complicated behavior may be expected.

Reprinted from THE PHYSICAL REVIEW D, Vol. 1, No. 12, 3296–3312, 15 June 1970

Momentum-Space Behavior of Integrals in Nonpolynomial Lagrangian Theories

ABDUS SALAM* AND J. STRATHDEE
International Atomic Energy Agency, International Centre for Theoretical Physics, Miramare, Trieste, Italy
(Received 13 October 1969)

Methods are developed for constructing momentum-space amplitudes corresponding to nonpolynomial nonderivative interactions of a real scalar field. The methods give rise to a supergraph technique and rules for writing down matrix elements very similar to Feynman techniques. The methods are not established rigorously; at several points the argument requires certain analytic properties of Feynman integrands which, though plausible, can only be demonstrated rigorously for the zero-mass case. Asymptotic behavior, both in spacelike and timelike directions, is discussed. Rough arguments are given that indicate that the singularity structure of the amplitudes is likely to be consistent with unitarity.

I. INTRODUCTION

IF it is to have any future, Lagrangian field theory must learn to cope with nonrenormalizable interactions. This becomes apparent when one examines what we currently believe are Lagrangians of physical interest.

1. These Lagrangians include the following.

(*a*) *Chiral $SU(2)\times SU(2)$ Lagrangians for strong interactions.* A typical example is Weinberg's Lagrangian for π mesons:

$$\mathfrak{L}=(\partial_\mu\phi)^2/(1+f\phi^2)^2.$$

(*b*) *Intermediate-boson-mediated weak Lagrangian.* An example is an intermediate neutral vector meson U_μ interacting with quarks Q. As is well known in Stückelberg's representation ($U_\mu=A_\mu+\kappa^{-1}\partial_\mu B$), $\mathfrak{L}_{\rm int}$ can be written in the typical form

$$\mathfrak{L}_{\rm int}=f\bar{Q}\gamma_\mu(1+\gamma_5)QA_\mu+m\bar{Q}(e^{i\gamma_5(f/\kappa)B}-1)Q.$$

(*c*) *The gravitational Lagrangian of Einstein* expressed in terms of the contravariant tensor $g^{\mu\nu}$

$$L=\kappa^{-2}(\sqrt{-g})g^{\mu\nu}(\Gamma_{\mu\rho}{}^\lambda\Gamma_{\nu\lambda}{}^\rho-\Gamma_{\mu\nu}{}^\lambda\Gamma_{\lambda\rho}{}^\rho),$$

where

$$\Gamma_{\mu\nu}{}^\lambda=\tfrac{1}{2}g^{\lambda\rho}(\partial_\mu g_{\nu\rho}+\partial_\nu g_{\mu\rho}-\partial_\rho g_{\mu\nu}).$$

The covariant components $g_{\mu\nu}$ which enter the expression for $g=\det g_{\alpha\beta}$ are expressed as a ratio of two polynomials in $g^{\mu\nu}$.

The interaction Lagrangians in all these theories are typically of a nonpolynomial form in field variables. These Lagrangians can be expanded in power series of the type

$$\mathfrak{L}_{\rm int}(\phi)=G\sum_n\frac{v(n)}{n!}(-\phi)^n. \tag{1.1}$$

(Here ϕ is a scalar field, and for simplicity we are ignoring derivatives.) The coefficients $v(n)$ are proportional to f^n, where f is a coupling constant.[1] All terms in such

* On leave of absence from Imperial College, London, England.

[1] In this paper we distinguish between the coupling constants G and f: G will be called the *major* coupling constant and f the *minor*. We shall be considering fixed order in G and all orders in f.

expansions with $n>4$ are nonrenormalizable. Thus, barring quantum electrodynamics, the rest of Lagrangian particle physics apparently needs a closer study with respect to nonrenormalizability.

2. Right from the very early days it was emphasized, particularly by Heisenberg, that the perturbation expansions of the S matrix in powers of $G^N f^r$ for the case of unrenormalizable theories suffer from two distinct (though related) difficulties.

(*a*) *Infinities.* The integrals in the theory become more and more infinite in each increasing order of perturbation. In each order one needs new counterterms containing *higher and higher derivatives of fields* if the conventional subtraction philosophy of renormalizable interactions is to be extended to these theories. (These infinite-order higher derivatives are likely to produce a *nonlocal* counter-term Lagrangian.)

(*b*) *Unacceptable high-energy behavior.* Even after a successful subtraction scheme has been carried out, the high-energy behavior (of the finite parts) of integrals is physically unacceptable. As external momenta become large, the dependence of these integrals on external momenta increases polynomially with the order of the approximation, unlike the case for renormalizable theories.

3. Of these two types of difficulties, the first—concerning the infinities of the integrals—has begun to be seriously investigated recently. Three types of approaches have been considered.

(*a*) *The conventional approach*, where the Feynman momentum-space integrals in each order $G^N f^r$ are considered as they stand and a consistent subtraction procedure defined.[2] It appears[3] that all three rigorous subtraction procedures used for renormalizable theories, i.e., (i) the Dyson-Salam method,[4] (ii) the Bogolubov-Parasiuk-Hepp method,[5] and (iii) the analytic renormalization method,[6] can be extended to nonrenormalizable theories. To our knowledge, the second technical problem of a systematic organization of the counterterms has not yet been examined for any of the theories, nor has the problem of physical interest posed by *finite* changes in the definition of renormalization constants. (Since there are an infinity of renormalization constants, such changes could reduce the predictive power of the theory to naught.)

(*b*) *The x-space approach of Efimov and Fradkin*[7,8] for theories with rational nonpolynomial Lagrangians. Formally one can write the Nth order approximation in the major coupling constant G in x space to a typical amplitude in the form of a divergent series:

$$F^N(x_1,x_2,\dots)=G^N\sum_{p,q\cdots}(f)^{p+q+\cdots}a_{pq\cdots}\times\Delta_F{}^p(x_1-x_2)\Delta_F{}^q(x_2-x_3)\cdots. \quad (1.2)$$

Efimov and Fradkin have described an elegant technique of carrying out Borel sums of such series *in the minor coupling parameter* f. These sums can be examined in the ultraviolet limit $(x_1-x_2)^2\to 0$, $(x_3-x_4)^2\to 0$, ... The important result of their investigation (extended in Ref. 8) is that if the Dyson index D of the rational Lagrangians is less than or equal to *four*—i.e., the same as that for renormalizable theories—[the Dyson index D is defined by the limit $L(\phi)_{\phi\to\infty}=\phi^D$], only a few types of Borel sums exhibit any ultraviolet infinities—again like the case for renormalizable theories. In particular, if the Dyson index D is less than two, none of the Borel sums (including those representing vacuum-to-vacuum transitions) is ultraviolet infinite. Thus if in some sense the Borel sums represent the physical amplitudes, all theories with $D<2$ (and for these theories the Lagrangian *must* be nonpolynomial) are *super-renormalizable.*[9]

This is a beautiful result. The important question to decide is to what extent the Borel sums represent the physical amplitudes. Do the p-space Fourier transforms of these x-space functions possess the requisite analyticity and unitarity properties? And, finally, is the high-energy behavior of these p-space Fourier transforms polynomially bounded, as it should be if physical amplitudes are being represented?

A detailed study of the Fourier transform of the Efimov-Fradkin two-point function in *second order of the major coupling constant* G has been made by Lee and Zumino,[10] who have concluded (with Efimov) that (i) the corresponding Borel sum does possess the requisite analyticity and unitarity properties, (ii) it is not polynomially bounded, and (iii) the well-known lack of uniqueness of Borel sums of divergent series is reflected in an arbitrariness of the amplitudes up to an entire function.

[2] The subtractions do not affect whatever causality and unitarity properties the perturbation expansion may have.

[3] K. Hepp, International Centre for Theoretical Physics, Trieste, Report No. IC/69/121 (unpublished).

[4] F. J. Dyson, Phys. Rev. **75**, 1736 (1949); Abdus Salam, *ibid.* **84**, 426 (1951). That this method carries through for nonrenormalizable theories has been shown in a set of basic papers by W. Zimmerman (unpublished).

[5] N. N. Bogolubov and O. S. Parasiuk, Acta Math. **97**, 227 (1957); K. Hepp, Commun. Math. Phys. **6**, 161 (1967).

[6] T. Gustafson, Arkiv. Mat. Astron. Fysik **34A**, No. 2 (1947); C. G. Bollini, J. J. Giambiagi, and A. Gonzalez Dominguez, Nuovo Cimento **31**, 550 (1964); E. R. Speer, J. Math. Phys. **9** 1404 (1968).

[7] G. V. Efimov, Zh. Eksperim. i Teor. Fiz. **44**, 2107 (1963) [Soviet Phys. JETP **17**, 1417 (1963)]; E. S. Fradkin, Nucl. Phys. **49**, 624 (1963); H. M. Fried, Nuovo Cimento **52A**, 1333 (1967); S. Okubo, Progr. Theoret. Phys. (Kyoto) **11**, 80 (1954) (this elegant paper was unknown to the authors until its existence was pointed out to them by Professor H. Lehmann); R. Arnowitt and S. Deser, Phys. Rev. **100**, 349 (1955).

[8] R. Delbourgo, Abdus Salam, and J. Strathdee, Phys. Rev. **187**, 1999 (1969). This paper will be referred to as I.

[9] As shown in Ref. 8, for the chiral π-meson Lagrangian the Dyson index is zero; for Einstein's gravity theory it equals *unity* [see R. Delbourgo, Abdus Salam, and J. Strathdee, Nuovo Cimento Letters **2**, 354 (1969)].

[10] B. W. Lee and B. Zumino Nucl. Phys. **B13**, 671 (1969).

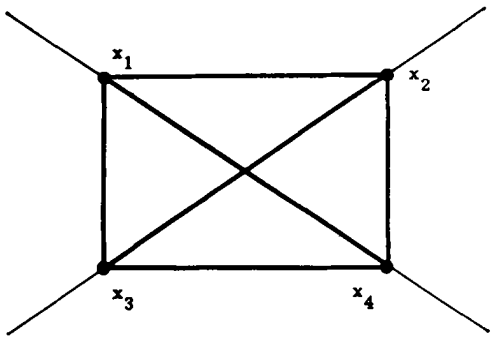


FIG. 1. Four-point supergraph. Heavy lines represent superpropagators each of which corresponds to the collection of functions Δ_F^n, $n=1, 2, 3\ldots$.

(c) *The p-space method.* Since it is the momentum-space Fourier transforms of the amplitude (1.2) which are the quantities of primary physical interest, it is valuable to have a summation method which works directly within p space. The present paper is devoted to the development of such a method, following a procedure first discussed in this context by Volkov[11] and which in its essentials goes back to a discussion (in the appropriate region of x and n) of the Fourier transform of $(-1/x^2)^n$ by Gel'fand and Shilov.[12] In particular we show the following.

(i) The amplitudes appear to possess the analyticity structure associated with the unitarity requirements.

(ii) The method immediately gives the asymptotic behavior for large values of external momenta; and, in particular, for the two-point amplitude in the second order in G^2 studied by Lee and Zumino, we reproduce their result very simply.

(iii) The discussion of ultraviolet infinities of Borel sums in x space is closely parallelled by a similar one in p space.

(iv) One can develop a graph technique of Feynman-like diagrams with *superlines* representing *superpropagators* $[\Delta_F(x)]^n$ replacing normal lines corresponding to Feynman propagators $\Delta_F(x)$. In p space, the closed-loop integrations for supergraphs can be performed with the help of Feynman's auxiliary parameters in exactly the same fashion as for conventional polynomial Lagrangians. Insofar as there is (essentially) just *one* supergraph in each order G^N, the topological analysis of supergraphs is simpler than Feynman diagrams for polynomial Lagrangians.

4. It would appear from the above that for non-polynomial interaction Lagrangians with index $D<2$, one can construct amplitudes with no ultraviolet infinities and which (if one can extrapolate from the limited experience so far) are likely to satisfy the correct analyticity and unitarity requirements (unitarity verified to each order in the major coupling constant G). There are two remaining problems:

(a) *The arbitrariness* in the amplitudes which the Borel-summation method in x space allows or, as we shall see, its weaker analog, which still exists when the p-space method is used. The problem is analogous to the problem of arbitrariness of finite renormalization constants in renormalizable theories.

(b) *The more serious problem* on nonpolynomial bounded high-energy behavior of the amplitudes. We believe that any inference in this respect on the basis of order-by-order calculations in powers of G is likely to be misleading, and a final verdict on the true asymptotic behavior of these theories can only be given after a summation of the series in the major coupling constant G has been performed. For renormalizable theories, as is well known, this summation has been carried out for certain sequences of graphs for the four-point function and for a number of production amplitudes. The result is the emergence of Regge behavior for large values of energy, unsuspected if one had only considered individual terms of the perturbation expansion. We believe, on the basis of certain indicative considerations, that a similar (drastic) change in the high-energy behavior also occurs in the present theories when the summation in G is carried out. The p-space method is extremely convenient for summing the supergraphs insofar as the analytical expressions for the supergraphs resemble those for conventional polynomial Lagrangian theories.

5. To make the plan of the present paper clear and to bring out the parallel sets of ideas involved in the x-space and the p-space methods, we set down here a brief and nonrigorous summary of the paper.

(A) *Supergraphs.* Consider

$$\mathcal{L}_{\rm int}(\phi)=G\sum_{n=0}^{\infty}\frac{v(n)}{n!}(-\phi)^n$$

[$v(n)$ contains the minor coupling parameter f^n. The factor $(-)^n$ is included for later convenience.]

It is easy to verify that the G^N contribution to an amplitude $F(x_1,\ldots,x_N)$ with E external line can be written as a sum of contributions from a set of *supergraphs* constructed as follows.

(a) Take N points $x_1, x_2, \ldots, x_N$.

(b) Join all points pair-wise with just one superline joining two distinct points (x_i,x_j); associate with this line a positive integer n_{ij}.

(c) For each line write the factor

$$(1/n_{ij}!)[\Delta_F(x_i-x_j)]^{n_{ij}}.$$

(d) For each point x_i write a vertex factor

$$v(\textstyle\sum_j n_{ij}+m_i).$$

Here m_i is the number of external lines impinging on the point x_i.

[11] M. K. Volkov, Ann. Phys. (N.Y.) 49, 202 (1968).
[12] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I.

(e) The contribution of the supergraph to the amplitude equals

$$F_{m_1 m_2 \cdots}(x_1, \ldots, x_N)$$

$$= G^N \sum_{n_{ij}} \prod_i v(\sum_j n_{ij} + m_i) \prod_{i<j} \frac{[\Delta_F(x_i - x_j)]^{n_{ij}}}{n_{ij}!} . \quad (1.3)$$

(f) To get the total contribution in order G^N, sum over all partitions, $[m_1, \ldots, m_N]$, of the external lines with m_i lines at the ith vertex such that

$$\sum m_i = E.$$

(g) In the above set of rules we omit all tadpole contributions (lines joining a point x_i to itself). This is justified if we consider instead of (1.1) a suitably Wick-ordered interaction. Figures 1 and 2 show a typical supergraph and a superline.

(h) It is clear that the Green's function

$$F_{m_1, m_2, m_3 \cdots}(x_1, \ldots, x_N),$$

with m_1, m_2 nonzero, is simply related to

$$F_{m_1 - 1, m_2 - 1, m_3, \cdots}.$$

For example,

$$\Delta(x_1 - x_2) F_{1,1,0,0,\cdots} = \left. \frac{\partial F_{0,0,0,\cdots}(\lambda)}{\partial \lambda} \right|_{\lambda = 1},$$

where $F(\lambda)$ on the right-hand side of (1.4) is obtained from (1.3) by replacing $\Delta(x_1 - x_2)$ by $\lambda \Delta(x_1 - x_2)$.

(*B*) *To illustrate* the x-space and p-space techniques, consider a simple example with $\mathfrak{L}_{\text{int}}$:

$$V(\phi) = \frac{G}{1 + f\phi} = G \sum_{n=0}^{\infty} (-f\phi)^n. \quad (1.4)$$

Here $v(n) = f^n n!$

(*a*) *The formal series expansion for amplitudes.* Formally an expectation value like

$$F(\Delta) = F_{00}(x_1, x_2) = \langle V(\phi(x_1)), V(\phi(x_2)) \rangle$$

equals the *divergent* series:

$$= G^2 \sum_{n=0}^{\infty} n! f^{2n} \Delta_F^n(x_1 - x_2). \quad (1.5)$$

We are interested in giving a meaning to this divergent series such that the Fourier transform

$$\tilde{F}(p^2) = \int F(\Delta) e^{ip \cdot x} d^4x \quad (1.6)$$

possesses correct analyticity and unitarity properties.

(*b*) *The Euclidicity postulate.* To do this consider the Symanzik region in p space ($p^2 < 0$). (When more than one external momentum p_i is involved, the Symanzik region is the region for which $p_i^2 \leqslant 0$, $p_i p_j \leqslant 0$.) Following Efimov, we can define the integral (1.6) by making a Wick rotation $x_0 \to i x_4$. For this region in p space, one therefore needs to consider $\Delta(x)$ for Euclidean x space only. [For a zero-mass field $\Delta(x) = -1/4\pi^2 x^2$ is real and positive.] For p-space regions outside the Symanzik region, we must appropriately analytically continue (1.6). [It cannot be emphasized strongly enough that for divergent series of the type (1.5), one is not starting by "proving" the validity of the Wick rotation. Rather, Euclidicity is a basic postulate—part of the process of defininining the theory. One *accepts* it for the Symanzik region; outside this region one makes an analytic continuation.]

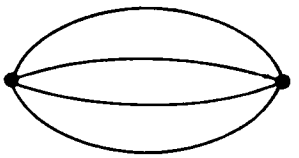


FIG. 2. Typical graph from the collection which is represented by a heavy line.

(*c*) *Borel summation.* To give meaning to the divergent sum $F(\Delta)$, use Borel transforms and write

$$F(\Delta) = \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\zeta} (f^2 \zeta \Delta)^n. \quad (1.7)$$

Here we have used

$$n! = \int_0^{\infty} \zeta^n e^{-\zeta} d\zeta.$$

(*d*) *The x-space method.* The x-space method consists of inverting integration and summation in (1.7) and writing it as

$$F(\Delta) = \int_0^{\infty} d\zeta \, e^{-\zeta} (1 - \zeta f^2 \Delta)^{-1}. \quad (1.8)$$

The expression (1.8) defines the amplitude $F(\Delta)$. At this stage we encounter our first problem in the x-space method; the integrand has a pole on the integration path at

$$\zeta = \frac{1}{f^2 \Delta} = \frac{4\pi^2 r^2}{f^2} \quad \text{for the case } m = 0, \; r^2 = \mathbf{x}^2 + x_4^2.$$

We must define how to go around this singularity.

One obvious answer is: Take the principal value (P.V.). This is because $F(\Delta)$ in the Symanzik region is a sum of real terms. The P.V. prescription for the integral representation (1.8) of $F(\Delta)$ will guarantee this. Lee and Zumino show that this is essentially the correct prescription, barring an arbitrariness (to be specified later) associated with functions like $\exp[1/(f^2 \Delta)]$

which possess an identically vanishing asymptotic expansion about the point $\Delta=0$, and which can be added to $f(\Delta)$ without affecting its representation in the form $\sum_{n=0}^{\infty} n!(f^2\Delta)^n$.

(*e*) One may now compute the Fourier transform, using $r=(-x^2)^{1/2}$,

$$\tilde{F}(p^2)=\int dx\, e^{ipx}\mathrm{P.V.}\int_0^\infty d\zeta\, e^{-\zeta}(1-\zeta f^2\Delta)^{-1}$$

$$=\frac{4\pi^2}{(-p^2)^{1/2}}\int_0^\infty dr\, r^2 J_1((-p^2)^{1/2}r)$$

$$\times\mathrm{P.V.}\int_0^\infty d\zeta\, e^{-\zeta}(1-f^2\zeta\Delta)^{-1} \quad (1.9)$$

for $p^2<0$ and continue direct to $0<p^2<m^2$. Inverting the order of integrations, we have

$$\tilde{F}(p^2)=\frac{4\pi^2}{(-p^2)^{1/2}}\int_0^\infty d\zeta\, e^{-\zeta}\int_0^\infty dt\, r^4 J_1((-p^2)^{1/2}r)$$

$$\times[r^2-\zeta(f/2\pi)^2]^{-1}.$$

This integral can be explicitly evaluated, and a continuation to timelike values of p^2 carried out, to demonstrate that $\tilde{F}(p^2)$ possesses the correct analyticity structure in the p^2 plane. The asymptotic behavior of $\tilde{F}(p^2)$ is

$$\tilde{F}(s)\to 1/(f^2s)^3, \quad s\to-\infty$$
$$\to\pm i\pi\exp(f^2s), \quad s\to+\infty\pm i0, \quad (1.10)$$

where $s=p^2$.

(*f*) *The p-space method.* Our procedure is different. It depends on Volkov's observation of the power of the Gel'fand-Shilov investigation of the Fourier transform of the generalized function $[\Delta(m=0)]^z=r^{-2z}$ in the range $0<\mathrm{Re}z<2$. [In Sec. II we consider the case $m\neq 0$. We have no exact expression for this case similar to (1.11) below, but the general considerations are parallel to those treated here.]

The crucial formula is

$$\Delta^z(x)=\frac{1}{(2\pi)^4}\int d^4p\, e^{-ipx}\frac{(-p^2)^{z-2}\pi(4\pi)^{2-2z}}{\sin\pi z\,\Gamma(z)\Gamma(z-1)},$$

$$0<\mathrm{Re}z<2. \quad (1.11)$$

To use this formula go back to the Borel sum (1.7) and employ a *Sommerfeld-Watson transformation* to convert the series into an integral of the form

$$F(\Delta)=\tfrac{1}{2}i\int_\Gamma\frac{dz}{\sin\pi z}\int d\zeta\, e^{-\zeta}(-\zeta f^2\Delta)^z, \quad (1.12)$$

with the contour Γ enclosing the positive real axis in the z plane. (The conditions for validity of employing this transformation are discussed in Sec. III.) Straighten the contour to lie along the imaginary axis with Rez constrained to lie in the range $0<\mathrm{Re}z<2$. Using the Gel'fand-Shilov formula to take the Fourier transform, we obtain

$$\tilde{F}(p^2)=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\frac{(-f^2)^z(-p^2)^{z-2}\Gamma(z+1)}{\sin\pi z\,\Gamma(z)\Gamma(z-1)}$$

$$+(2\pi)^4\delta(p), \quad (1.13)$$

where $0<\alpha<2$. [The term $\delta(p)$ corresponds to a graph which contains no internal line.]

In the above passage from (1.5) to (1.13), we have changed orders of summation and integration repeatedly. The justification is provided in Sec. III. Here we are only concerned with a rapid exposition of the p-space method whose chief ingredients are the Sommerfeld-Watson transformation (1.12) and the Fourier transformation (1.11).

(*g*) Formula (1.13) is our master formula. By closing the contour along the left, one can immediately obtain the asymptotic behavior of $\tilde{F}(p^2)$ for $p^2\to-\infty$ [and, in particular, the Lee-Zumino result (1.10)]. As in Regge-pole theory, the rightmost pole of the integrand gives the leading contribution to the asymptotic behavior.

(*h*) The principal value ambiguity of the x-space method noted in (d), which does not seem to arise explicitly in this treatment, has a weaker counterpart when we take into account the appearance of the negative sign in front of Δ in $(-\Delta)^z$ in the Sommerfeld-Watson transform. To see this more explicitly, introduce a multiplier λ in front of Δ; thus,

$$F(\lambda\Delta)=\tfrac{1}{2}i\int\frac{dz}{\sin\pi z}\int d\zeta\, e^{-\zeta}(-\zeta\lambda f^2\Delta)^z. \quad (1.14)$$

We must interpret the limit $\lambda\to+1$ by a *real* average of the values $(-\lambda)^z=e^{i\pi z}$ and $(-\lambda)^z=e^{i\pi z}$, obtaining finally

$$F(\Delta)=\int dz\left(\frac{1}{\tan\pi z}+b\right)\Gamma(z+1)(f^2\Delta)^z \quad (1.15)$$

with b an arbitrary real constant. The ambiguity introduced by $(-1)^z=e^{i\pi z}=e^{-i\pi z}$ we shall call the *signature ambiguity* of superpropagators. (Note that this method admits of an arbitrariness only up to a constant, and not up to an entire function of p^2, as in the x-space method.)

(*j*) In Paper I we showed that a superficial count indicates that the number of distinct types of ultraviolet infinities depends on the Dyson index D $[\lim_{\phi\to\infty}V(\phi)=\phi^D]$ of the Lagrangian. This number is finite if $D\leqslant 4$. The result was proved by considering $x\to 0$ (x spacelike) behavior of Borel sums (1.7). For example, explicitly when $V(\phi)=G/(1+f\phi)$, with

$D=-1$, $F(\Delta)$ is given by

$$F(\Delta)=G^2\,\mathrm{P.V.}\int_0^\infty d\zeta\, e^{-\zeta}x^2\left[x^2-\zeta\left(\frac{f}{2\pi}\right)^2\right]^{-1}$$

for zero-mass fields ϕ. This expression is finite in the limit $x\to 0$. For the interaction Lagrangian $V(\phi)=\phi^5/(1+f\phi)$ with $D=4$, however, we recover the ultraviolet infinities since the corresponding expression for $F(\Delta)$ is

$$F(\Delta)=\Delta^5\int d\zeta\, e^{-\zeta}(1-f^2\zeta\Delta)^{-1}.$$

The question arises: Where in the p-space method is there an indication of a Dyson index? The answer, as we shall see in Sec. III, is that it is the Gel'fand requirement $0<\mathrm{Re}z<2$ for the unique definition of the Fourier transform of $\Delta^z(x)$ which forces us to distinguish between Lagrangians like $V(\phi)=1/(1+f\phi)$ and $V(\phi)=\phi^5/(1+f\phi)$. In order to give a precise meaning to $F(\Delta)$ for the latter, we are constrained to write it in the form

$$F(\Delta)=\sum_{n=2}^{4} n!\Delta^n+\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty} dz\,\frac{(-f^2\zeta\Delta)^z}{\sin\pi z}e^{-\zeta}d\zeta,\ (1<\alpha<2).$$

The terms which appear in the sum $\sum_{n=2}^{4} n!\Delta^n$ are just the ones which give rise to ultraviolet infinities.

(*k*) *Higher orders.* The great beauty of the p-space method lies in the similarity of the p-space expressions for supergraphs and normal Feynman diagrams.

One can introduce Feynman's auxiliary parameters and carry out the loop integrations. As we shall see below, the result is an elegant expression for the supergraph contribution as a weighted average integral of contributions of conventional graphs. The utility of such an expression is twofold. (i) The sums of supergraphs in different orders of G closely resemble the sums for conventional graphs, and the methods previously discussed by Polkinghorne, Federbush,[13] and others for carrying through the summation can be taken over. (ii) The discontinuity formulas of Cutkosky—and the proof of the unitarity relations using such formulas—follow the conventional lines.

For the zero-mass case, the integral expression for the Nth-order supergraph is the following.

Associate with each superline a four-momentum vector q_{ij}. The Sommerfeld-Watson transform of (1.3) in p space equals

$$\bar{F}(p)=G^N\left(\prod_{i<j}\int dz_{ij}\rho(z_{ij})\int d^4q_{ij}(-q_{ij}^2)^{z_{ij}-2}\right)\times\prod_k\delta^4\Big(p_k+\sum_{l\neq k}q_{kl}\Big).\quad(1.16)$$

Here $\rho(z_{ij})$ is the product of the vertex factors $v(\sum_{j\neq i}z_{ij}+m_i)$, the factors $1/\sin\pi z_{ij}$, and the factors $1/[\sin\pi z_{ij}\Gamma(z_{ij})\Gamma(z_{ij}-1)]$ for each superline. The p_i's are the momenta carried by the external lines at the ith vertex, and the δ functions express conservation of energy and momentum.

Introduce Feynman's auxiliary parameters, using the integral representation[14]

$$(-q^2)^{z-2}=\frac{1}{\pi\Gamma(2-z)}\int_0^\infty d\alpha\,\alpha^{1-z}e^{\alpha q^2}.\quad(1.17)$$

One may now carry through the d^4q integrations in

$$I(p,\alpha_{ij})=\int(\exp\textstyle\sum\alpha_{ij}q_{ij}^2)\prod_k\delta^4\Big(p_k+\sum_{l\neq k}q_{kl}\Big)\Big]^N\prod_{i<j}d^4q_{ij}.\quad(1.18)$$

The result is identical to the case as if we were dealing with *normal* Feynman graphs with $F=\frac{1}{2}N(N-1)$ internal lines rather than supergraphs. [This is because $I(p_i,\alpha_{ij})$ is not z_{ij} dependent.] Such normal graphs we shall call *skeleton graphs*. The evaluation of the functions $I(p_i,\alpha_{ij})$ for the skeleton graphs can easily be carried through using the methods of Chisholm[15]; the final expression for the amplitude $\bar{F}(p_i)$ reads

$$\bar{F}(p)=\prod_{i<j}\int dz_{ij}\rho'(z_{ij})\int d\alpha_{ij}\alpha_{ij}^{1-z_{ij}}I(p,\alpha_{ij}),\quad(1.19)$$

where ρ' differs from ρ by the factors

$$\prod_{i<j}\frac{1}{\pi\Gamma(2-z_{ij})}.$$

The result for the N-point function evaluated in order G^N can therefore be stated thus: Draw a normal Feynman graph with internal lines joining all the N points pair wise. We shall call such graphs skeleton graphs. Introduce Feynman parameters; the result of performing loop integrations in skeleton diagrams is the standard Chisholm expression $I(p,\alpha_{ij})$. Multiply this by the factors $(\alpha_{ij})^{1-z_{ij}}$ and the weight function $\rho'(z_{ij})$; integrate over Feynman parameters α_{ij} and the Sommerfeld-Watson parameters z_{ij}. One obtains the supergraph contribution.

(*m*) *Finiteness of the supergraphs.* One may examine the supergraph integrals for ultraviolet infinities. It is easy to see that a superficial power count would indicate that for an N-point function, with $F=\frac{1}{2}N(N-1)$ internal lines and $l=F-N+1$ loop momenta, the supergraphs have no ultraviolet infinities provided each superpropagator contributes a factor falling like $(q^2)^{-2}$. It is crucial to remember that with our Euclidicity

[13] J. C. Polkinghorne, J. Math. Phys. 4, 503 (1963); P. G. Federbush and M. T. Grisaru, Ann. Phys. (N. Y.) 22, 263 (1963); 22, 299 (1963).

[14] For $z=1$ we recover Feynman's formula for normal propagators.

[15] J. R. S. Chisholm, Proc. Cambridge Phil. Soc. 48, 300 (1952).

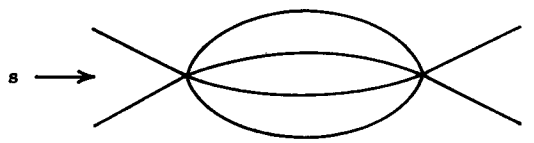


FIG. 3. Contribution to the s-wave scattering amplitude, $\bar{F}(s)\sim\exp[(\text{const})s]$, $s>0$.

postulate it is the asymptotic behavior for spacelike q^2 for the superpropagators which is relevant here. This also leads us to stress once again that it would be a mistake to evaluate supergraphs in each order in the major coupling parameter G^N for the Symanzik region, to continue the external momenta to the physical region and *then* sum the series in G^N. *One must sum in G^N first* (as indeed has been done for the series in the minor coupling constant f) *and then continue the sum to the physical region.*

(n) As stated earlier, the G^2 approximation to the two-point amplitude in a typical nonpolynomial theory behaves asymptotically, as illustrated in Figs. 3 and 4. Thus, to order G^2, a simple evaluation for form factors (Fig. 4) immediately yields physically sensible results[16] $[\bar{F}(t)\approx G^2/t^2]$, while the same approximation in the timelike region $s>0$ gives, for the four-particle scattering amplitude, a physically unacceptable behavior $[\bar{F}(s)\propto G^2e^{\alpha s}]$. It is clear that before rejecting nonrenormalizable theories on the grounds of unacceptability of their predictions in the lowest-order calculation in the timelike region of external momenta, one must first carry out a summation of a chain of diagrams. This crucial problem is being studied.

II. SUPERPROPAGATOR

As we have been stressing in Sec. I, the basic generalized function, in terms of which all the Green's functions are ultimately to be expressed, is Δ^z, where z denotes a complex number in the strip $0<\text{Re}z<2$, and Δ is the usual propagator for a free scalar field of mass m,

$$\langle T(\phi(x)\phi(0))\rangle=\Delta(x^2-i0,m^2). \tag{2.1}$$

The function $\Delta(x^2,m^2)$ is analytic in the x^2 plane cut from 0 to $+\infty$. Explicitly,

$$\Delta(x^2,m^2)=mK_1(m\sqrt{(-x^2)})/4\pi^2\sqrt{(-x^2)}, \tag{2.2}$$

where K_1 denotes the modified Hankel (or Macdonald) function. The function Δ is real and positive on the negative real axis. It has no zeros in the finite x^2 plane, and its behavior near $x^2=0$ and $x^2=\infty$ is given by

$$\begin{aligned}\Delta(x^2,m^2)&\sim -1/x^2, \quad x^2\to 0\\ &\sim(-x^2)^{-3/4}e^{-m\sqrt{(-x^2)}}, \quad |x^2|\to\infty.\end{aligned} \tag{2.3}$$

The generalized function Δ^z, which we shall call the superpropagator, is well defined provided one can find a space of test functions $f(x)$ over which the integral

$$(\Delta^z,f)=\int d^4x\,\Delta^z f(x) \tag{2.4}$$

is convergent and satisfies the appropriate continuity conditions. This integral can certainly be defined for $f(x)\in\mathfrak{D}$, the space of infinitely differentiable functions with bounded support, provided z lies in the strip $0<\text{Re}z<2$. (Presumably it can be extended to larger spaces but we have not examined this problem.) Following the standard procedure[12] for defining a generalized function that corresponds to an ordinary function with an algebraic singularity, we can define (Δ^z,f) outside the strip $0<\text{Re}z<2$ by means of analytic continuation. The result is an analytic function of z with simple poles at the integers $z=2, 3, 4, \ldots$.

The Fourier transform $D(p^2,z)$ of the superpropagator must, like the latter, be an analytic function of z. It is defined on the segment $\text{Im}z=0$, $0<\text{Re}z<2$ by the classical integral

$$D(p^2,z)=\frac{1}{i}\int d^4x\,e^{ipx}\Delta^z, \tag{2.5}$$

which converges absolutely for $p^2<0$. In fact, one can perform a Wick rotation of the x_0 contour and replace the Minkowskian integral (2.5) by an equivalent Euclidean one which reduces to the form

$$\begin{aligned}&D(p^2,z)\\ &=\frac{4\pi^2}{(-p^2)^{1/2}}\int_0^\infty dr\,r^2J_1((-p^2)^{1/2}r)\Delta(-r^2,m^2)^z\end{aligned} \tag{2.6}$$

after performing the angular integrations. It is clear that (2.6) converges for a wider range of z than does (2.5): $0<\text{Re}z<2$, $-\infty<\text{Im}z<\infty$. The analytic continuation of $D(p^2,z)$ outside this strip will be considered below. It will be shown that like the functional (Δ^z,f), it has poles at the integers $z=2, 3, 4, \ldots$.

For the zero-mass case we can express the integral (2.6) in terms of elementary functions and so perform the analytic continuation explicitly. The result is

$$D_0(p^2,z)=\pi(4\pi)^{2-2z}\frac{(-p^2)^{z-2}}{\sin\pi z\,\Gamma(z)\Gamma(z-1)}, \tag{2.7}$$

which clearly exhibits the poles at $z=2, 3, 4, \ldots$. (It shows in addition the rather unexpected feature of *zeros* at $z=0, -1, -2, \ldots$. We have not yet been able

[16] That a typical form factor falls so fast is a welcome result. It is perhaps not a surprising result, since it has already been pointed out in the literature that falling form factors are most likely to be consequences of the existence of multiparticle intermediate states. See, for example, J. S. Ball and F. Zachariasen, Phys. Rev. 170, 1541 (1968); D. Amati, L. Caneschi, and R. Jengo, Nuovo Cimento 98, 783 (1968); A. O. Barut and H. Kleinert, Phys. Rev. 161, 1464 (1967); C. Fronsdal, *ibid.* 171, 1811 (1968).

to prove that this is true also of the massive superpropagator.)

The asymptotic behavior of the superpropagator both in p^2 and in z is of particular importance. For $|p^2|\to\infty$ with z fixed, we shall assume that $D(p^2,z)$ can be approximated by $D_0(p^2,z)$. For $|z|\to\infty$ with p^2 fixed, the situation is less clear. The behavior of the zero-mass superpropagator

$$D_0\sim(-p^2)^{z-2}|-z|^{-2\mathrm{Re}z-2},\quad |\arg z|<\pi \tag{2.8}$$

may or may not provide a useful guide to the massive case.

Consider the structure of $D(p^2,z)$ in the finite z plane with $p^2<0$. The integral representation (2.6) is valid only in the strip $0<\mathrm{Re}z<2$. In order to continue to the right of $\mathrm{Re}z=2$, we must modify the behavior of the integrand of (2.6) at $r=0$. We propose to subtract and add the first N terms of a Maclaurin expansion of J_1. Indeed, if we write

$$\frac{J_1((-p^2)^{1/2}r)}{(-p^2)^{1/2}r}=\frac{1}{2}\sum_{k=0}^{N-1}\frac{(\frac{1}{4}p^2r^2)^k}{k!(k+1)!}+R_N(p^2r^2), \tag{2.9}$$

then (2.6) takes the form

$$D(p^2,z)=2\pi^2\sum_{k=0}^{N-1}\frac{(\frac{1}{4}p^2)^k}{k!(k+1)!}\int_0^\infty dr\, r^{2k+3}\Delta(-r^2,m^2)^z$$

$$+4\pi^2\int_0^\infty dr\, r^3R_N(p^2r^2)\Delta(-r^2,m^2)^z. \tag{2.10}$$

The term involving R_N is easily shown to converge in the extended strip

$$0<\mathrm{Re}z<N+2 \tag{2.11}$$

since near $r=0$ we have $R_N(p^2r^2)\sim r^{2N}$. The other terms in (2.10) will of course exhibit poles when continued out of the strip $0<\mathrm{Re}z<2$. (This can be shown by the method of Gel'fand and Shilov.) The important point here is the fact that the residues of the poles at $z=2, 3, 4$ are *polynomials* in p^2. (This is simply a reflection of the well-known fact that the ultraviolet divergences manifest themselves in the coefficients of a polynomial.) By increasing N indefinitely, we can thus prove that $D(p^2,z)$ with $p^2<0$ is analytic in the half-plane $\mathrm{Re}z>0$ except at the points $z=2, 3, 4, \ldots$ where it has simple poles, the residues of which are polynomials in p^2 (of order $z-2$). The structure in the half-plane $\mathrm{Re}z<0$ is more difficult to unravel and we have not attempted this.

Consider now the structure of $D(p^2,z)$ in the p^2 plane. It is trivial but helpful to continue into the strip $0<\mathrm{Re}(p^2)^{1/2}<m\,\mathrm{Re}z$ by expressing the integral (2.6) in the modified form

$$D(p^2,z)=\frac{4\pi^2}{(p^2)^{1/2}}\int_0^\infty dr\, r^2I_1((p^2)^{1/2}r)\Delta(-r^2,m^2)^z. \tag{2.12}$$

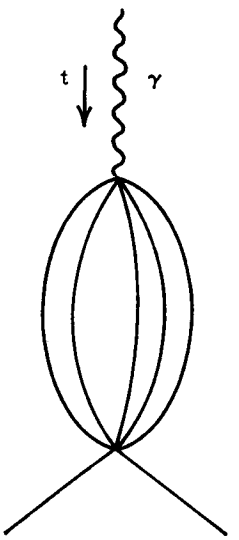


FIG. 4. Contribution to the electromagnetic form factors $\tilde{F}(t)$ $\sim 1/t^2$, $t<0$.

It is the convergence or lack of it at the upper limit which controls the analyticity in p^2. We shall be able to make analytic continuations by displacing the r contour in (2.12). The integrand has no singularities in the finite plane, and for large values of $|r|$ it can be approximated by

$$r^{\frac{1}{2}(1-z)}e^{-(mz-\sqrt{p^2})r},$$

assuming $\mathrm{Re}(p^2)^{1/2}>0$. Let us keep z fixed in the strip $1<\mathrm{Re}z<2$. It is clear that we can rotate the r contour through the angle θ without affecting the value of the integral, provided we maintain the condition

$$\mathrm{Re}[(mz-\sqrt{p^2})e^{i\theta}]>0,$$

or, otherwise expressed,

$$|\arg(mz-\sqrt{p^2})+\theta|<\tfrac{1}{2}\pi. \tag{2.13}$$

This follows from the absence of singularities of Δ^z in the r plane apart from the branch point at $r=0$ where the convergence does not depend on $\arg(r)$. Starting with the original contour ,$\theta=0$, we have analyticity in the half-plane

(I) $\quad -\tfrac{1}{2}\pi<\arg(mz-\sqrt{p^2})<\tfrac{1}{2}\pi.$

Increasing θ continuously to $+\tfrac{1}{2}\pi$ rotates the convergence domain into

(II) $\quad -\pi<\arg(mz-\sqrt{p^2})<0,$

while decreasing θ to $-\tfrac{1}{2}\pi$ rotates the domain into

(III) $\quad 0<\arg(mz-\sqrt{p^2})<\pi.$

The regions (I)–(III) so obtained can be pictured in the plane of $\sqrt{p^2}$ as in Fig. 5. The dashed lines indicate boundaries between the regions. Thus it appears that the integral (2.12) defines a function which is analytic in the half-plane $\mathrm{Re}(\sqrt{p^2})>0$ except at the point $\sqrt{p^2}$

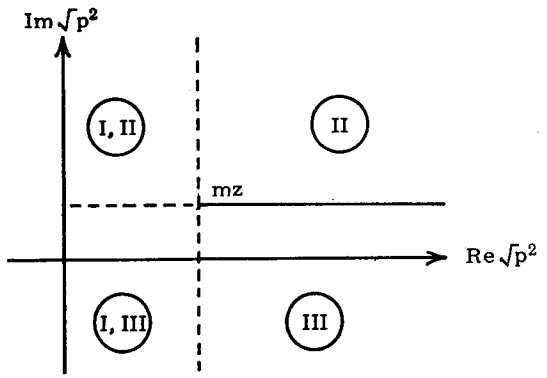


FIG. 5. Structure of the superpropagator in the complex $\sqrt{p^2}$ plane. The solid line represents a branch cut and the dashed lines mark out regions in which the superpropagator is represented by distinct contour integrals.

$=mz$, where it presumably has a branch point. The discontinuity across the branch cut is given by $D_{II}-D_{III}$, where D_{II} and D_{III} are defined by the contours with $\theta=\frac{1}{2}\pi-0$ and $\theta=-\frac{1}{2}\pi+0$, respectively. Equivalently,

$$D_{II}(p^2,z)=\frac{4\pi^2}{(p^2)^{1/2}}\int_0^\infty du\,u^2J_1((p^2)^{1/2}u)\times\left(\frac{im}{8\pi}\frac{H_1^{(2)}(mu)}{u}\right)^z, \tag{2.14}$$

$$D_{III}(p^2,z)=\frac{4\pi^2}{(p^2)^{1/2}}\int_0^\infty du\,u^2J_1((p^2)^{1/2}u)\times\left(-\frac{im}{8\pi}\frac{H_1^{(1)}(mu)}{u}\right)^z.$$

These formulas are valid for $1<\mathrm{Re}z<2$. If we add and subtract the Maclaurin terms as in (2.10), they can be extended in an obvious way into the half-plane $\mathrm{Re}z>2$. Presumably their range of validity can be extended down to $\mathrm{Re}z=0$ without difficulty (at $z=1$ the branch point should reduce to a simple pole).

To summarize, from the integral representation (2.6) and its modifications (2.10), (2.12), and (2.14) we deduce that the superpropagator $D(p^2,z)$ viewed as a function of two complex variables has the singularities: (a) fixed poles at $z=2, 3, \ldots$, the residues of which are polynomials of order $z-2$ in p^2; (b) a singular surface $p^2=z^2m^2$ which manifests itself as a branch point in the p^2 plane except when $z=1$, in which case it becomes a simple pole. The discontinuity across the branch cut is a regular function of z—at least for $\mathrm{Re}z>0$. Whenever z is an integer the singularity surface $p^2=z^2m^2$ corresponds to normal thresholds as implied by unitarity.

We have no clear idea of the behavior of the superpropagator for $\mathrm{Re}z\leqslant 0$, except, of course, in the zero-mass case where it is analytic with zeros at $z=0, -1, -2, \ldots$.

For large values of $|p^2|$ it is presumably adequate to approximate $D(p^2,z)$ by the zero-mass form (2.7). For large values of $|z|$ it may or may not be possible to use (2.8).

Finally, let us consider an alternative regularization of the superpropagator Δ^z which can be used in the zero-mass case. (This regularization will be referred to in Sec. III.) Introduce the regularizing parameter a to define

$$\Delta_{\mathrm{reg}}(-r^2,0)=\frac{1}{4\pi^2}\frac{1}{r^2+a^2},\quad a>0 \tag{2.15}$$

which has no singularity at $r=0$. Substituting this form into (2.6), we find the corresponding momentum-space superpropagator

$$D_{\mathrm{reg}}(p^2,z)=\frac{(2\pi)^{2-2z}}{(-p^2)^{1/2}}\int_0^\infty dr\,\frac{r^2J_1((-p^2)^{1/2}r)}{(r^2+a^2)^z}=\frac{2(4\pi^2)^{1-z}}{\Gamma(z)}\left(\frac{(-p^2)^{1/2}}{2a}\right)^{z-2}K_{z-2}(a(-p^2)^{1/2}), \tag{2.16}$$

and, as was to be expected, the fixed poles at $z=2, 3, \ldots$ have disappeared. In order to see what happens in the limit $\alpha\to 0$ let us assume that z is not an integer and replace K_{z-2} by its series expansion. The result is

$$D_{\mathrm{reg}}(p^2,z)=\frac{\pi}{\sin\pi z}(4\pi)^{2-2z}\frac{(-p^2)^{z-2}}{\Gamma(z)\Gamma(z-1)}\Bigg[1-\frac{1}{z-1}\frac{a^2p^2}{4}+\frac{1}{2z(z-1)}\left(\frac{a^2p^2}{4}\right)^2-\cdots-a^{4-2z}\frac{\Gamma(z-1)}{\Gamma(3-z)}\left(-\frac{p^2}{4}\right)^{2-z}-a^{6-2z}\frac{\Gamma(z-1)}{\Gamma(4-z)}\left(-\frac{p^2}{4}\right)^{3-z}-\cdots\Bigg]. \tag{2.17}$$

In the limit $a\to 0$ every term except the first vanishes if $\mathrm{Re}z<2$. In this way the superpropagator (2.7) can be obtained as the limit of a regularized function.

If, on the other hand, we were to take $2<\mathrm{Re}z<3$, then the limit would be singular:

$$D_{\mathrm{reg}}(p^2,z)\to D_0(p^2,z)-(4\pi^2)^{1-z}a^{4-2z}/(z-1)(z-2). \tag{2.18}$$

This means that contour integrals involving $D_{\mathrm{reg}}(p^2,z)$ must be moved to the left of $\mathrm{Re}z=2$ *before* the regularization is removed.

III. REGULARIZED PERTURBATION SERIES

Consider now the problem of developing perturbation expansions in the nonpolynomial interaction $V(\phi)$. As

discussed in Sec. I, we could begin formally by expanding V in powers of ϕ and then, for each power ϕ^n, develop the usual series. Let us therefore suppose that the interaction Hamiltonian, considered as a function of the complex variable ϕ, is analytic in some neighborhood of $\phi=0$ so that we can write

$$V(\phi)=\sum_n (v_n/n!)(-\phi)^n, \quad |\phi|<R. \tag{3.1}$$

It is convenient for our purposes to consider, instead of the usual many-body Green's functions, the equivalent set of amplitudes

$$(-)^{\Sigma m}F_{m_1\cdots m_N}(\Delta)=\langle T(V^{(m_1)}(\phi_1)\cdots V^{(m_N)}(\phi_N))\rangle, \tag{3.2}$$

where $\phi_1=\phi(x_1)$, ... and $V^{(m)}(\phi)=\partial^m V(\phi)/\partial\phi^m$. If we substitute the power series (3.1) into (3.2) then, after some straightforward manipulations, we obtain the series

$$F_{m_1\cdots m_N}(\Delta)\simeq\sum_{n_{ij}} v(m_1+\sum_j n_{1j})\cdots v(m_N+\sum_j n_{Nj}) \times\prod_{i<j}[\Delta(x_i-x_j)]^{n_{ij}}/n_{ij}!, \tag{3.3}$$

where the indices i, j, run from 1 to N and the $n_{ij}=n_{ji}$ from 0 to ∞. We shall suppose that $n_{ii}=0$ since the factors $\Delta(0)^{n_{ii}}$ produce no more than a rather simple renormalization of coupling strengths which we can regard as having been done already.[17]

For the interactions we shall be considering, the series (3.3) generally diverge. Thus, for example, corresponding to the interaction (1.4) we have $v(n)=Gf^n n!$ and so in the simplest case $N=2$, $m_1=m_2=0$,

$$F_{00}(\Delta)\simeq G^2\sum_{n=0}^{\infty} f^{2n}n!\Delta^n,$$

which clearly diverges. However, we can easily regard such series as asymptotic expansions and direct our attention to the problem of defining the amplitudes which are so represented. As a first step towards such a definition, let us consider the Borel sum of (3.3),

$$F_{m_1\cdots m_N}{}^B(\Delta)=\int_0^\infty\left(\prod_{i<j}d\zeta_{ij}e^{-\zeta_{ij}}\sum_{n_{ij}=0}^{\infty}\right)v(m_1+\sum n_{ij})\cdots \prod_{i<j}\frac{[\zeta_{ij}\Delta(x_i-x_j)]^{n_{ij}}}{n_{ij}!}. \tag{3.4}$$

The summations over n_{ij} are by this method—or some generalization[18] of it—made to converge so that we are left with the problem of integrating over the ζ_{ij}. The ζ integrals may well be undefined since the analytic function in the integrand can develop singularities on the path of integration. These can always be avoided by distorting the ζ contours or by giving the Δ's suitable imaginary parts.

The precise details of the method used to define the functions $F^B(\Delta)$ are not important since we are using the Borel integrals only at an intermediate stage of our program. The essential step will be the replacing of the sums over n_{ij} in (3.4) by contour integrals over z_{ij}. To do this we need to interpolate the expansion coefficients v_n by an analytic function $v(z)$. The existence of such an interpolating function is assured by the condition[19]

$$\int_0^\infty d\phi|\phi^{-\alpha-1}V(\phi)|<\infty \tag{3.5}$$

for some range of α. Indeed, if (3.5) is satisfied then $V(\phi)$ may be represented by the Mellin integral

$$V(\phi)=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\frac{v(z)}{\Gamma(z+1)}\phi^z, \tag{3.6}$$

where $v(z)$ is an analytic function defined by the integral

$$v(z)=\frac{1}{\Gamma(-z)}\int_0^\infty d\phi\,\phi^{-z-1}V(\phi), \quad \mathrm{Re}z=\alpha. \tag{3.7}$$

In order to compare the integral representation (3.6) with the power series (3.1) we have only to perform the inverse Watson-Sommerfeld transformation. It appears that α is constrained by the minimum value n_0 of n appearing in the sum (3.1), i.e.,

$$n_1<\alpha<n_0, \tag{3.8}$$

where n_0 and n_1 are determined by the limiting behavior

$$\begin{aligned} V(\phi)&\sim\phi^{n_0}, \quad \phi\to 0\\ &\sim\phi^{n_1}, \quad \phi\to\infty. \end{aligned} \tag{3.9}$$

[We shall be assuming later that $V(\phi)$ can be expanded in powers of $1/\phi$ for $|\phi|>R_1$. In this expansion the lowest power to appear is $-n_1$.]

To avoid the notational obscurities consequent upon the use of general formulas like (3.4), we devote the remainder of this section to a detailed treatment of the second-order amplitudes,

$$F_{m_1m_2}{}^B(\Delta)=\int_0^\infty d\zeta\, e^{-\zeta}\sum_{n=0}^{\infty}v(m_1+n)v(m_2+n)\frac{(\zeta\Delta)^n}{(n!)^2}. \tag{3.10}$$

Formal generalizations for the higher-order amplitudes can usually be made without difficulty. New features connected with the problem of defining products of superpropagators may appear in the higher orders; about these we can draw only tentative conclusions. These problems will be presented in Sec. IV.

[17] R. Delbourgo and K. Koller (unpublished).

[18] Examples of such generalizations are given in Refs. 7 and 8.

[19] E. C. Titchmarsh, *Theory of Fourier Integrals*, 2nd ed. (Oxford U. P., Oxford, England, 1967), p. 46.

It will prove convenient in the following to have an auxiliary complex variable λ at our disposal. Let us therefore consider the functions $F^B(\lambda\Delta)$, bearing in mind that we shall take a limit $\lambda \to 1$ at the end. Let us now replace the summation over n in (3.10) by a contour integral using the $v(z)$ of (3.7), analytically continued, to interpolate the coefficients v_n,

$$F_{m_1m_2}{}^B(\lambda\Delta)=\int_0^\infty d\zeta\, e^{-\zeta}(\tfrac{1}{2}i)\int_\Gamma \frac{dz}{\sin\pi z} \times \frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)^2}(-\lambda)^z(\zeta\Delta)^z, \quad (3.11)$$

where Γ denotes a contour coming from $+\infty$ which encircles the integers $z=M, M+1, M+2, \ldots$, in the negative sense, and returns to ∞. The non-negative integer M is fixed as the lowest power of Δ appearing in the sum (3.10). It is defined by

$$M=\max(n_0-m_1, n_0-m_2, 0). \quad (3.12)$$

Let us now open the contour Γ after the fashion of Watson and Sommerfeld. This is possible because of the postulated existence of the Mellin representation for $V(\phi)$. The factor $\Gamma(z+1)^{-2}$ produces a strong damping effect for $|z| \to \infty$, $|\arg z| < \frac{1}{2}\pi$. Once the contour has been opened, however, the factor

$$\Gamma(z+1)^{-2} \sim \exp(\pi|\mathrm{Im}z|)$$

acts to weaken convergence. Hence it is advisable at this stage to interchange the ζ and z integrals and perform the ζ integration to obtain

$$F_{m_1m_2}{}^B(\lambda\Delta) = \tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty} \frac{dz}{\sin\pi z}\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)}(-\lambda)^z\Delta^z, \quad (3.13)$$

where α lies in the range

$$M_1<\alpha<M; \quad (3.14)$$

here M_1 denotes the position of the singularity of the integrand nearest on the left to $z=M$. This will generally be a pole at $z=M-1$ due to the factor $(\sin\pi z)^{-1}$. However, if $M=1$ or 0, this will be a pole from one or both of the factors $v(z+m_1)$ and $v(z+m_2)$. Let us defer these questions until we come to consider the construction of asymptotic series for the momentum-space amplitudes.

In formula (3.13) we have thus obtained an integral representation of the Mellin type involving the superpropagator Δ^z. If $M_1<2$, then this integral defines unambiguously the generalized functions $F^B(\lambda\Delta)$. If, on the other hand, $M_1 \geqslant 2$, then it will be necessary to translate the contour to the left of the line $\mathrm{Re}z=2$ and, in so doing, pick up the poles at $z=2, 3, \ldots[M_1]$. The part of the generalized function defined by the new contour is unambiguous but the separated terms, involving $\Delta^2, \Delta^3, \ldots, \Delta^{[M_1]}$, are not. They carry the usual ultraviolet divergences.

A simple illustration of this effect can be given in the zero-mass case using the regularized superpropagator $D_{\mathrm{reg}}(p^2,z)$ in place of Δ^z in (3.13). This propagator, given by (2.16), has no z poles and in the strip $0<\mathrm{Re}z<2$ it reduces to the correct form (2.7) in the limit when the regularization is removed. However, if $\mathrm{Re}z>2$ then singular terms appear when the regularization is removed. That is, the Fourier transform of $F^B(\lambda\Delta)$ is well defined when the regularization is removed only if the contour is contained in the strip $0<\mathrm{Re}z<2$.

Assuming now that $0<\alpha<2$, we can immediately write down the Fourier transform of $F^B(\lambda\Delta)$. It is given by

$$\tilde{F}_{m_1m_2}{}^B(p^2,\lambda)=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z} \times \frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)}(-\lambda)^z D(p^2,z). \quad (3.15)$$

At this point we must remark that the continuation to $z=0$ of $D(p^2,z)$ must be treated carefully. For the zero-mass case the explicit formula (2.7) gives

$$D(p^2,0)=0.$$

On the other hand, we should expect the Fourier transform of $\Delta^0=1$ to be $(2\pi)^4\delta^4(p)$. This means that if, in the integral representation (3.13), the contour lies to the left of $z=0$, i.e., if $M=0$, then before defining the Fourier transform we must translate the contour to the right of $z=0$ where we can use (3.15) and, in compensation, add the term

$$v(m_1)v(m_2)(2\pi)^4\delta^4(p), \quad (3.16)$$

which is the Fourier transform of the contribution of the pole at $z=0$. Clearly (3.16) corresponds to a disconnected graph. The contour method apparently picks out the connected graphs only.

Let us now consider the problem of constructing an asymptotic series to represent the function $\tilde{F}^B(p^2,\lambda)$. To simplify the discussion let us suppose that $M=0$ or 1. (If $M \geqslant 2$ the necessary modifications can be made without difficulty.) For large spacelike p^2, we shall use the zero-mass approximation for $D(p^2,z)$ and write (3.15) in the form

$$\tilde{F}_{m_1m_2}{}^B(p^2,\lambda) \simeq \tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)^2} \times \frac{z(-\lambda)^z(-p^2)^{z-2}}{\sin\pi z\,\Gamma(z-1)}(4\pi)^{2-2z}$$

$$=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)^2}\frac{z}{\pi} \times \Gamma(2-z)(-\lambda)^z(-p^2)^{z-2}(4\pi)^{2-2z}. \quad (3.17)$$

Our aim, in order to get a series in inverse powers of p^2, is to collapse the contour onto the negative real axis—a mirror image of the conventional Watson-Sommerfeld contour on the positive real axis. By our assumption that $V(\phi)$ can be expanded in inverse powers of ϕ for sufficiently large $|\phi|$, it follows that $v(z)/\Gamma(z)$ is damped with sufficient strength for the inverse Watson-Sommerfeld transformation to go through in the half-plane $\mathrm{Re}z<0$. The only factor in (3.17) which hinders this operation is $\Gamma(2-z)$, which explodes. However, this can be removed by the Borel trick. That is, we can write

$$\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda)\sim\int_0^\infty\frac{d\zeta}{\zeta}e^{-\zeta}(\tfrac{1}{2}i)\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\times\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)^2}\frac{z}{\pi}(-\lambda)^z\left(-\frac{p^2}{\zeta}\right)^{z-2}(4\pi)^{2-2z}. \quad (3.18)$$

The dominant term as $p^2\to\infty$ is going to come from the pole or dipole of $v(z+m_1)v(z+m_2)$ furthest to the right. For simplicity let us suppose that the poles of $v(z)$ occur at negative integer values of z. This happens in many cases of interest and it is a consequence of requiring that $V(\phi)$ has a Laurent expansion at $\phi=\infty$. Suppose

$$V(\phi)=\sum_n\frac{u(n)}{n!}(-\phi)^{-n}.$$

Comparing this with the integral representation (3.6), we find

$$v(z)=-\frac{\Gamma(z)}{\Gamma(-z)}u(-z),$$

so that, in particular,

$$\frac{v(z+m)}{\Gamma(z+1)}=-(z+m)(z+m-1)\cdots(z+1)\frac{u(-z-m)}{\Gamma(1-z-m)}, \quad (3.19)$$

which vanishes for $z=-1, -2, \ldots, -m$. It will vanish also for

$$z=-m-1, \ldots, -m-n_2+1,$$

where n_2 denotes the lowest power of ϕ occurring in the regular part of the Laurent expansion of $V(1/\phi)$, i.e., $u(n)=0$ for $n=0, 1, \ldots, n_2-1$. The poles of the integrand of (3.18) therefore occur at

$$z=-N-1, -N-2, \ldots,$$

where N is given by

$$N=\max(m_1, m_2, m_1+n_2-1, m_2+n_2-1). \quad (3.20)$$

Hence we can write

$$\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda)\sim\int_0^\infty\frac{d\zeta}{\zeta}e^{-\zeta}\sum_{n=N+1}^{\infty}(-)^n\frac{(n-1)!}{(n-m_1-1)!}\times\frac{u(n-m_1)}{(n-m_1)!}\frac{(n-1)!}{(n-m_2-1)!}\frac{u(n-m_2)}{(n-m_2)!}\times(-\lambda)^n\frac{n}{\pi}\left(-\frac{\zeta}{p^2}\right)^{n+2}(4\pi)^{2(n-1)}. \quad (3.21)$$

Finally, to obtain the asymptotic[20] series we interchange the ζ integration with the summation. We obtain the result

$$\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda)\simeq\frac{1}{\pi}\sum_{n=N+1}^{\infty}\lambda^n\times\frac{[(n-1)!]^2u(n-m_1)u(n-m_2)}{(n-m_1-1)!(n-m_1)!(n-m_2-1)!(n-m_2)!}\times n(n+1)!\left(-\frac{4\pi^2}{p^2}\right)^{n+2}. \quad (3.22)$$

There are two important aspects of this formula. Firstly, the leading term in $\tilde{F}^B$ turns out to be $(-p^2)^{-N-2}$ where N is given by (3.20). [If $M\geqslant 2$ we must include additional terms corresponding to the poles at $z=1, 2, \ldots, M_1$. These are $(1/p^2), \ln p^2, \ldots (p^2)^{M_1-2}\ln p^2$.] Secondly, the asymptotic series (3.22) is *single valued* in λ.

The functions $\tilde{F}^B(p^2,\lambda)$ defined by the integral (3.15) are not single valued in λ. Generally they have a logarithmic singularity at $\lambda=0$. This can be seen if we collapse the contour of (3.15) onto the positive real axis and pick up the residues of the *dipoles* at $z=2, 3, \ldots$. Let us use again the zero-mass expression for the superpropagator to write

$$\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda)=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dz}{\sin\pi z}\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)}(-\lambda)^z\frac{(-p^2)^{z-2}}{\sin\pi z\Gamma(z)\Gamma(z-1)}$$
$$=\frac{v(1+m_1)v(1+m_2)}{p^2}\lambda+\frac{1}{\pi}\sum_{n=2}^{\infty}\frac{\partial}{\partial z}\left[\frac{v(z+m_1)v(z+m_2)(-\lambda)^z(-p^2)^{z-2}}{\Gamma(z+1)\Gamma(z)\Gamma(z-1)}\right]_{z=n}$$
$$=\frac{v(1+m_1)v(1+m_2)}{p^2}\lambda+\frac{1}{\pi}\sum_{n=2}^{\infty}\frac{v(n+m_1)v(n+m_2)(-\lambda)^n(-p^2)^{n-2}}{n!(n-1)!(n-2)!}\times\left[\frac{v'(n+m_1)}{v(n+m_1)}+\frac{v'(n+m_2)}{v(n+m_2)}-\psi(n+1)-\psi(n)-\psi(n-1)+\ln(-\lambda)+\ln(-p^2)\right], \quad (3.23)$$

[20] We shall not attempt to prove that the series (3.22) is indeed asymptotic. Presumably one could ensure this by imposing appropriate conditions on $V(\phi)$.

where $\psi(z)=\Gamma'(z)/\Gamma(z)$. That is, we can write

$$\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda)=A_{m_1m_2}(p^2,\lambda)+\ln(-\lambda)B_{m_1m_2}(p^2,\lambda), \quad (3.24)$$

where A and B are entire functions of λ. Moreover, B is an entire function of p^2 as well.

That $B(p^2,\lambda)$ is an entire function of p^2 can be seen from the fact that the discontinuity of $\tilde{F}^B$ is an entire function of λ and must therefore come from the function $A(p^2,\lambda)$. Considering the massive case, we can use the property that the absorptive part of the superpropagator, $D(p^2,z)$, vanishes for $p^2<(mz)^2$ to evaluate the absorptive part of $\tilde{F}^B$ by translating the contour in (3.15) to the right of $\mathrm{Re}z=(\sqrt{p^2})/m$ for given $\sqrt{p^2}>0$. Only the poles separated in this way can contribute to the discontinuity, which is therefore given by

$$\mathrm{disc}\tilde{F}_{m_1m_2}{}^{B}(p^2,\lambda) = \sum_{n=1}^{[\sqrt{p^2}/m]} \frac{v_{m_1+n}v_{m_2+n}}{n!}\lambda^n \,\mathrm{disc}D(p^2,n), \quad (3.25)$$

which is just a polynomial in λ.

The form (3.24) which has been derived for the zero-mass case is probably true in general. Its validity depends only on the feasibility of the inverse Watson-Sommerfeld transformation together with the fact that $D(p^2,z)$ has simple poles at the integers $z=2, 3, \ldots$. These combine with the zeros of $\sin\pi z$ to make dipoles.

The dispersive part of $\tilde{F}^B$ is certainly not an integral function of λ and we shall have to adopt some definition of the limit $\lambda\to 1$. It is at this point that a basic uncertainty enters the program. In the absence of guidelines we can interpret $\lim_{\lambda\to 1}(-\lambda)^z$ by an average of the terms

$$e^{i(2k+1)\pi z}, \quad k=0,\pm1,\pm2,\ldots.$$

That is, we should write

$$\tilde{F}_{m_1m_2}(p^2)=\sum_k a_k\tilde{F}_{m_1m_2}{}^{B}(p^2,-e^{i(2k+1)\pi}) \quad (3.26)$$

with arbitrary complex parameters a_k. Substituting the form (3.24), this reads

$$\tilde{F}_{m_1m_2}(p^2)=(\sum_k a_k)A_{m_1m_2}(p^2,1) + i\pi(\sum_k (2k+1)a_k)B_{m_1m_2}(p^2,1), \quad (3.27)$$

so that there are really only two arbitrary constants.

There is one very important constraint to be imposed. That is unitarity. The imaginary part of $\tilde{F}(p^2)$ should be given by (3.25) with $\lambda=1$, and it should vanish for $p^2<0$. This gives us the conditions

$$\sum a_k=1 \quad \text{and} \quad i\pi\sum(2k+1)a_k=b, \quad (3.28)$$

where b is real. Thus it appears that the Fourier transform of the generalized function $F(\Delta)$ may contain one arbitrary parameter b. Possibly we could take $b=0$. This corresponds to the choice advocated in earlier references (Efimov[7] and Paper I[8]) and it may receive justification when we are able to impose the unitarity requirement in higher orders.

The final choice of amplitude $\tilde{F}(p^2)$ represented by (3.27) and (3.28) can be expressed by the integral

$$\tilde{F}_{m_1m_2}(p^2)=\tfrac{1}{2}i\int_{\alpha-i\infty}^{\alpha+i\infty} dz\,\frac{v(z+m_1)v(z+m_2)}{\Gamma(z+1)}D(p^2,z) \times\left(\frac{1}{\tan\pi z}+b\right). \quad (3.29)$$

IV. HIGHER ORDERS AND FEYNMANIZATION OF SUPERGRAPHS

Corresponding to a diagram with N vertices which are pairwise connected by superpropagators there exists the momentum-space amplitude

$$\int dx_1\cdots dx_N e^{i\Sigma px}\prod_{i<j}\Delta(x_i-x_j)^{z_{ij}} =(2\pi)^4\delta(\textstyle\sum p)D(p_1\cdots p_N;z_{12}\cdots) \quad (4.1)$$

depending upon $N-1$ independent momenta and $\frac{1}{2}N(N-1)$ independent complex parameters z_{ij}. The singularities of this integral occur on the various light cones $(x_i-x_j)^2=0$. An over-all convergence condition can be obtained by considering the behavior of the integrand when all components x_i approach 0 simultaneously. This gives the over-all singularity

$$x^{-2\Sigma\mathrm{Re}z_{ij}}=x^{-N(N-1)\mathrm{Re}z}$$

if we assume for simplicity that $\mathrm{Re}z_{ij}$ is the same for every z_{ij}. This singularity is compensated by $4(N-1)$ integrations. Hence we have superficial convergence if

$$N(N-1)\,\mathrm{Re}z<4(N-1),$$

i.e.,

$$\mathrm{Re}z<4/N. \quad (4.2)$$

An equivalent representation of the amplitude (4.1) is given by the momentum-space integral

$$D(p_1\cdots p_N;z_{12}\cdots)=\int\prod_{\text{loops}}(dk)\prod_{i<j}D(q_{ij}{}^2,z_{ij}), \quad (4.3)$$

where $D(q^2,z)$ denotes the superpropagator of Sec. II. The $\frac{1}{2}(N(N-1)$ momenta q_{ij} associated with the superlines are expressed in the usual way by linear combinations of the loop momenta $k_1, \ldots, k_l$ and the external momenta $p_1\cdots p_N$. The convergence of (4.3) can be justified by the same power-counting arguments as were used above. Using the asymptotic form $D(q^2,z)\sim(-q^2)^{z-2}$, one arrives again at the condition (4.2).

The problem of analyzing higher-order contributions is a very standardized one. For each N there is one and

only one skeleton graph. This graph is obtained by joining the vertices in pairs—one line to each pair. The resulting diagram has $\frac{1}{2}N(N-1)$ lines and $\frac{1}{2}(N-1)\times(N-2)$ loops. The amplitude which corresponds to this diagram would be highly divergent if ordinary bare propagators were associated with the lines. It is the possibility of taking $\mathrm{Re}z_{ij}$ sufficiently small which makes the amplitude converge when superpropagators are associated with the lines. The analytic function defined by this convergent integral can then be continued outside the original domain $\mathrm{Re}z_{ij}<4/N$ exactly as was done for the superpropagator itself.

In this way one is led to define the higher-order momentum-space amplitudes by the multiple contour integral

$$\tilde{F}_{m_1\cdots m_N}{}^{B}(p_1\cdots p_N;\lambda_{12}\cdots)$$
$$=\int\prod_{\text{lines}}\left(\tfrac{1}{2}i\frac{dz_{ij}}{\sin\pi z_{ij}}\frac{(-\lambda_{ij})^{z_{ij}}}{\Gamma(z_{ij}+1)}\right)$$
$$\times\prod_{\text{vertices}}v(m_k+\sum_l z_{kl})D(p_1\cdots p_N;z_{12}\cdots),\quad(4.4)$$

where the z contours lie initially in the strip $0<\mathrm{Re}z_{ij}<4/N$. The auxiliary parameters λ_{ij} must then be set equal to $+1$, and it is at this stage that some ambiguity can enter the problem. The general procedure should be to define the true amplitude as a linear combination of the possible limits $\lambda_{ij}\to 1$, i.e.,

$$\tilde{F}_{m_1\cdots m_N}(p_1\cdots p_N)$$
$$=\sum_{\nu_{12}\cdots}a_{\nu_{12}\cdots}{}^{N}\tilde{F}_{m_1\cdots m_N}{}^{B}(p_1\cdots p_N;-e^{i\nu_{12}\pi}\cdots),\quad(4.5)$$

where each ν_{ij} takes the values $+1$ and -1. The coefficients $a_{\nu_{12}\cdots}{}^{N}$ are to be chosen consistently with unitarity but are otherwise arbitrary. Substituting the representation (4.4) into (4.5) one obtains the form

$$\tilde{F}_{m_1\cdots m_N}(p_1\cdots p_N)=\int\prod_{\text{lines}}\left(\tfrac{1}{2}i\frac{dz_{ij}}{\sin\pi z_{ij}}\frac{1}{\Gamma(z_{ij}+1)}\right)$$
$$\times\prod_{\text{vertices}}v(m_k+\sum_l z_{kl})D(p_1\cdots p_N;z_{12}\cdots)$$
$$\times a^{N}(z_{12}\cdots),\quad(4.6)$$

where $a^{N}(z_{12}\cdots)$ denotes an entire function defined by the sum

$$a^{N}(z_{12}\cdots)=\sum_{\nu_{12}\cdots=\pm1}a_{\nu_{12}\cdots}{}^{N}\exp(i\pi\sum_{i<j}\nu_{ij}z_{ij}).\quad(4.7)$$

It is necessary to investigate in what way this "ambiguity function" (which resembles the signature ambiguity in Regge theory) is constrained by the requirements of unitarity.

The unitarity problem is of course an extremely intricate one and so we shall confine the discussion to a conjecture about normal thresholds. To this end let the N vertices be divided into two sets, $1, 2, \ldots, M$ and $1', 2', \ldots, M'$ $(M+M'=N)$. If the amplitude $D(p_1\cdots p_N;z_{ij})$, considered as a function of the variable $(p_1+\cdots+p_M)^2=(p_{1'}+\cdots+p_{M'})^2$ has a branch point at

$$(p_1+\cdots+p_M)^2=m^2(\sum_{ii'}z_{ii'})^2,\quad(4.8)$$

and if the discontinuity across the associated cut is given by the integral

$$\int(\prod dq_{ij'})\delta(p_1+\cdots+p_M-\sum q)$$
$$\times D(p_1+\sum q_{1i'},\ldots;z_{ij})\prod_{ij'}\theta(q_{ij'})\,\mathrm{disc}D(q_{ij'}{}^2,z_{ij'})$$
$$\times D(p_{1'}+\sum q_{1'i},\ldots;z_{i'j'}),\quad(4.9)$$

where the variable $q_{ij'}$ denotes the four-momentum carried from vertex i in the first set to vertex j' in the second, we would have a situation for supergraphs similar to the Landau-Cutkosky discontinuity formulas for normal Feynman graphs. The plausibility of (4.9) can be seen when we consider that the discontinuity (4.9) is a regular function of the $z_{ij'}$ which vanishes when the real part of $\sum z_{ij'}$ is taken sufficiently large. This follows from the properties, established in Sec. II, of the superpropagator $D(q^2,z)$.

If the expression (4.9) for the discontinuity of $D(p_1\cdots p_N;z_{ij})$ is used in conjunction with the representation (4.4), it is possible to give the discontinuity of $\tilde{F}^{B}$ in the form

$$\mathrm{disc}\tilde{F}_{m_1\cdots m_N}{}^{B}(p_1\cdots p_N;\lambda_{ij})=\int\prod_{\text{lines}}\left(\tfrac{1}{2}i\frac{dz_{ij}}{\sin\pi z_{ij}}\frac{(-\lambda_{ij})^{z_{ij}}}{\Gamma(z_{ij}+1)}\right)\prod_{\text{vertices}}v(m_k+\sum_{l\neq k}z_{kl})\,\mathrm{disc}D(p_1\cdots p_N;z_{ij})$$
$$=\int\prod_{\text{lines}}\left(\tfrac{1}{2}i\frac{dz_{ij}}{\sin\pi z_{ij}}\frac{(-\lambda_{ij})^{z_{ij}}}{\Gamma(z_{ij}+1)}\right)\prod_{\text{vertices}}v(m_k+\sum_{l\neq k}z_{kl})\int\prod(dq_{ij'})\delta(p_1+\cdots+p_M-\sum q_{ij'})$$
$$\times D(p_1+\sum q_{ij},\ldots;z_{ij})\prod_{ij'}\theta(q_{ij'})\,\mathrm{disc}D(q_{ij'}{}^2;z_{ij'})D(p_{1'}+\sum q_{1'i},\ldots;z_{i'j'}).\quad(4.10)$$

The appearance of this formula can be simplified considerably by translating to the right the contours $z_{ij'}$ corresponding to particles exchanged between the two sets. In the course of this translation simple poles due to the zeros of $\sin\pi z_{ij'}$ will be crossed and their residues must be collected. Ultimately, when the contours have been pushed far enough, the contribution of these contours to the discontinuity will vanish. The discontinuity is thereby expressed—for given $(p_1+\cdots+p_M)^2$—by the residues of the *finite* set of poles at $z_{ij'}=n_{ij'}$ which have been crossed. The result is

$$\begin{aligned}\mathrm{disc}\bar{F}_{m_1\cdots m_N}{}^{B}(p_1\cdots p_N;\lambda_{ij})&=\sum_{n_{ij'}}\int\prod(dq_{ij'}\theta(q_{ij'}))\delta(\textstyle\sum q_{ij'}-p_1-\cdots-p_M)\\&\times\bar{F}_{m_1+\Sigma n_{1j'},\ldots}{}^{B}(p_1+\textstyle\sum q_{1j'},\ldots,p_M+\sum q_{Mj'};\lambda_{ij})\\&\times\prod_{ij'}\lambda_{ij'}{}^{n_{ij'}}\Omega_{n_{ij'}}(q_{ij'}{}^2)\\&\times\bar{F}_{m_{1'}+\Sigma n_{1'i},\ldots}{}^{B}(p_{1'}+\textstyle\sum q_{1'i},\ldots,\lambda_{i'j'}),\end{aligned}\tag{4.11}$$

where $\Omega_n(q^2)$ denotes the n-particle phase volume.

A similar factorization of the discontinuity into the products of lower-order amplitudes will be obtained for the true amplitude (4.6), provided the entire function a^N factorizes according to

$$\lim_{z_{ij'}\to n_{ij'}}a^N(z_{12}\cdots)=a^M(z_{ij})a^{M'}(z_{i'j'})(-)^{\Sigma n_{ij'}}.\tag{4.12}$$

(For the two-point function discussed in Secs. I and III, $a^2(z)=\cos\pi z+b\sin\pi z$.) One possible form for general a^N which satisfies (4.12) is

$$a^N(z_{ij})=\prod_{i,j=1}^{N}a^2(z_{ij}),$$

though this may not be the most general one. This is a strong result and would imply that there is just one arbitrary constant b in the whole theory. A result similar to this but not as strong has been claimed by Efimov, who shows on the basis of unitarity that, in his recent formulation[21] of the theory, there is just one arbitrary function $b(s)$ associated with superpropagators.

Another method of attack on the unitarity problem which may give more insight is to eliminate the loop momenta from (4.3) in favor of a set of Feynman parameters. As will be seen below, this method has the advantage of making a sharper separation between the factors which depend on the details of the interaction and the kinematical factors which are common to all interactions. In fact, the momentum-space amplitude in Nth order can be expressed in the form

$$\bar{F}_{m_1\cdots m_N}(p_1\cdots p_N)=\int_0^\infty\prod_{\text{lines}}d\alpha_{ij}\mathbf{F}_{m_1\cdots m_N}(\alpha_{12}\cdots)\times\mathbf{C}(\alpha)^{-2}\exp\left(\frac{\mathbf{D}(\alpha,p)}{\mathbf{C}(\alpha)}\right),\tag{4.13}$$

where $\mathbf{C}(\alpha)$ and $\mathbf{D}(\alpha,p)$ are functions which are completely determined by the structure of the skeleton graph, which, for these considerations, will always be taken as the set of N vertices pairwise connected by $\frac{1}{2}N(N-1)$ lines. The functions $\mathbf{C}$ and $\mathbf{D}$ are those defined, for example, in Ref. 22 with the stipulation that zero-mass bare propagators be used in the definition. The function $\mathbf{F}_{m_1\cdots m_N}(\alpha_{12}\ldots)$ contains the dynamical information and also the ambiguities. It will be defined in the following.

The derivation of the integral representation (4.13) proceeds in the following way. Firstly, since the superpropagator $D(q^2,z)$ is analytic in the q^2 half-plane $\mathrm{Re}q^2<\mathrm{Re}(mz)^2$ and is bounded there by a power, it can be expressed as a Laplace transform

$$D(q^2,z)=\int_0^\infty d\alpha\,\tilde{D}(\alpha,z)e^{\alpha q^2},\quad \mathrm{Re}q^2<\mathrm{Re}(mz)^2.\tag{4.14}$$

The new amplitude $\tilde{D}$, the α representative of the superpropagator, is obtained by inverting this integral. For $\alpha>0$,

$$\begin{aligned}\tilde{D}(\alpha,z)&=\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}dq^2e^{-\alpha q^2}D(q^2,z),\quad \beta<\mathrm{Re}(mz)^2\\&=\frac{1}{2\pi i}\int_{(mz)^2}^{\infty}dq^2e^{-\alpha q^2}\,\mathrm{disc}D(q^2,z),\end{aligned}\tag{4.15}$$

where the latter form is obtained by collapsing the contour onto the cut which extends from $(mz)^2$ to $+\infty$ in the q^2 plane. The fixed poles at $z=2, 3, 4, \ldots$ contained in $D(q^2,z)$ are absent from its discontinuity and therefore also from the new amplitude $\tilde{D}$. It is clear that $\tilde{D}$, considered as a function of complex α, is analytic in the half-plane $\mathrm{Re}\alpha>0$. In general, there is a singularity at $\alpha=0$, where

$$\tilde{D}(\alpha,z)\sim\frac{\alpha^{1-s}}{\Gamma(1-z)},\quad \alpha\to 0.\tag{4.16}$$

For $|\alpha|\to\infty$, $|\arg\alpha|<\frac{1}{2}\pi$, one finds

$$\tilde{D}(\alpha,z)\sim e^{-\alpha(mz)^2},\quad \mathrm{Re}z^2>0\tag{4.17}$$

[21] G. V. Efimov, Kiev Report Nos. ITF 68-52, ITF 68-54, and ITF 68-55 (unpublished).

[22] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U. P., Cambridge, England, 1966), Chap. II.

provided $m>0$. For the zero-mass case,

$$\tilde{D}(\alpha,z)=\frac{(4\pi)^{2-2z}}{\Gamma(z)}\alpha^{1-z} \tag{4.18}$$

exactly.

The Laplace representation (4.14) is easily generalized to higher forms like (4.3). One can do this formally by substituting an integral like (4.14) for each factor in (4.3) and then exchanging the loop integrals with the α integrals. One obtains

$$D(p_1\cdots p_N;z_{12}\cdots)$$
$$=\int_0^\infty \prod_{\text{lines}}[d\alpha_{ij}\tilde{D}(\alpha_{ij},z_{ij})]\int \prod_{\text{loops}}(dk)e^{\Sigma\alpha q^2}. \tag{4.19}$$

This exchanging of the integrals can of course be justified only in the Symanzik region of the external momenta and only after a Wick rotation has been applied so as to make all the loop integrals Euclidean, $q_{ij}^2\leqslant 0$. Under these circumstances the loop integrals in (4.19) can be performed explicitly to give

$$\int \prod_{\text{loops}}(dk)e^{\Sigma\alpha_{ij}q_{ij}^2}=\mathbf{C}(\alpha)^{-2}\exp\left(\frac{\mathbf{D}(\alpha,p)}{\mathbf{C}(\alpha)}\right), \tag{4.20}$$

where $\mathbf{C}(\alpha)$ denotes a homogeneous polynomial of degree l while $\mathbf{D}(\alpha,p)$ is linear in the invariants p_ip_j and homogeneous of degree $l+1$ in $\alpha_{12}\cdots$. The degree l is equal to the number of loops in the skeleton graph which in this case is given by $l=\frac{1}{2}(N-1)(N-2)$.

To compute an amplitude, one must multiply (4.19) by the appropriate vertex factors and integrate over the z_{ij}, i.e.,

$$\tilde{F}_{m_1\cdots m_N}{}^B(p_1\cdots p_N;\lambda_{12}\cdots)$$
$$=\int \prod_{\text{lines}}\left(\frac{dz_{ij}}{2\pi i}\Gamma(-z_{ij})(-\lambda_{ij})^{z_{ij}}\right)$$
$$\times \prod_{\text{vertices}} v(m_k+\sum_l z_{kl})D(p_1\cdots p_N;z_{12}\cdots)$$
$$=\int_0^\infty \prod_{\text{lines}}(d\alpha_{ij})\mathbf{F}_{m_1\cdots m_N}{}^B(\alpha_{12}\cdots;\lambda_{12}\cdots)$$
$$\times\mathbf{C}(\alpha)^{-2}\exp\left(\frac{\mathbf{D}(\alpha,p)}{\mathbf{C}(\alpha)}\right), \tag{4.21}$$

where the new amplitude $\mathbf{F}^B$ is defined by

$$\mathbf{F}_{m_1\cdots m_N}{}^B(\alpha_{12}\cdots;\lambda_{12}\cdots)$$
$$=\int \prod_{\text{lines}}\left(\frac{dz_{ij}}{2\pi i}\Gamma(-z_{ij})(-\lambda_{ij})^{z_{ij}}\tilde{D}(\alpha_{ij},z_{ij})\right)$$
$$\times \prod_{\text{vertices}} v(m_k+\sum_l z_{kl}). \tag{4.22}$$

There is no need to leave (4.22) in the form of a contour integral. One could collapse all of the z contours onto the positive axes and collect the residues of the poles there. The resulting sum, which represents the amplitude within some hypercircle of convergence, is given by

$$\mathbf{F}_{m_1\cdots m_N}{}^B(\alpha_{12}\cdots;\lambda_{12}\cdots)$$
$$=\sum_{n_{ij}} \prod_{\text{lines}}(\lambda_{ij}{}^{n_{ij}}\tilde{D}(\alpha_{ij},n_{ij})1/n_{ij}!)$$
$$\times \prod_{\text{vertices}} v(m_k+\sum_l n_{kl}). \tag{4.23}$$

For the class of theories considered in this paper, the series (4.23) converges for sufficiently large $|\alpha|$ and defines an analytic function of the Feynman parameters. In general, this function has singularities, some of which move onto the positive real axes when $\lambda_{ij}\to+1$. This phenomenon necessitates a distortion of the α contours in the integral (4.9). In order to define a sensible momentum-space amplitude, it will be necessary to take an average of the limits $-\lambda_{ij}\to\exp(i\pi\nu_{ij})$ with $\nu_{ij}=\pm1$ exactly as was done above.

To illustrate this, consider once more the second-order vacuum graphs corresponding to the interaction

$$v_n=Gf^n n!$$

in the zero-mass approximation. For this case, (4.11) reads

$$\mathbf{F}_{00}{}^B(\alpha,\lambda)=G^2\sum_n \lambda^n\frac{\alpha^{1-n}}{(n-1)!\,n!}f^{2n}(n!)^2$$
$$=G^2\frac{\lambda f^2\alpha^2}{(\alpha-\lambda f^2)^2}.$$

Corresponding to the definitions adopted in Sec. III, one can define the "true" amplitude by the limit

$$\mathbf{F}_{00}(\alpha)=\tfrac{1}{2}(1+ib)\mathbf{F}_{00}{}^B(\alpha,-e^{i\pi})+\tfrac{1}{2}(1-ib)\mathbf{F}_{00}{}^B(\alpha,-e^{i\pi})$$
$$=G^2f^2\alpha^2[\text{P.V.}(\alpha-f^2)^{-2}-\pi b\delta'(\alpha-f^2)],$$

where P.V. ( ) denotes the principal value and b is real. The momentum-space amplitude which corresponds to this is given by (4.13) with $\mathbf{C}(\alpha)=1$ and $\mathbf{D}(\alpha,p)=\alpha p^2$:

$$\tilde{F}_{00}(p^2)$$
$$=G^2f^2\left[\text{P.V.}\int_0^\infty d\alpha\frac{\alpha^2e^{\alpha p^2}}{(\alpha-f^2)^2}+\pi bf^2(f^2p^2+2)e^{f^2p^2}\right]$$
$$=G^2\left[p^2\left(\frac{\partial}{\partial p^2}\right)^3+2\left(\frac{\partial}{\partial p^2}\right)^2\right]e^{f^2p^2}E^*(-f^2p^2)$$
$$+\pi bG^2f^4(f^2p^2+2)e^{f^2p^2},$$

where $E^*(z)$ denotes the exponential integral function[23] cut from 0 to $-\infty$. The term containing $E^*(-f^2p^2)$

[23] Erdélyi, Magnus, Oberhettinger, and Tricomi, *Higher Transcendental Functions* (McGraw-Hill New York, 1953), Vol. II, p. 147.

behaves asymptotically like $(1/p^2)^3$ provided $|\arg p^2| > 0$. The term containing b, an entire function of p^2, dominates the asymptotic behavior if $|\arg(p^2)| \leqslant \frac{1}{2}\pi$. This may be a good reason for taking $b=0$.

In general, the result of this averaging of limits will be an integral like (4.13), with $\mathbf{F}$ given by

$$\mathbf{F}_{m_1\cdots m_N}(\alpha_{12}\cdots) = \sum_{\nu_{12}\cdots=\pm 1} a_{\nu_{12}\cdots}\mathbf{F}_{m_1\cdots m_N}{}^{B}(\alpha_{12}\cdots; -e^{i\nu_{12}\pi}\cdots), \quad (4.24)$$

a limit which must be interpreted in the sense of generalized functions. That is, $\mathbf{F}(\alpha)$ will have prescribed singularities such as $\text{P.V.}(\alpha-\alpha_0)^{-n}$ or $\delta(\alpha-\alpha_0)$ on the integration contours.

Before discussing the unitarity problem it will be useful to have yet another representation at one's disposal. In the power series (4.23), one can substitute for the α representative of the superpropagator the form

$$\tilde{D}(\alpha,n) = \int_0^\infty d\kappa^2\theta(\kappa^2-(nm)^2)e^{-\alpha\kappa^2}\Omega_n(\kappa^2),$$

where $\Omega_n(\kappa^2)$ denotes the n-particle phase volume. One obtains in this way, after making an interchange of the κ_{ij}^2 integrals with the n_{ij} sums,

$$\mathbf{F}_{m_1\cdots m_N}(\alpha_{12}\cdots) = \int_0^\infty \prod_{\text{lines}} (d\kappa_{ij}^2)\sigma_{m_1\cdots m_N}(\kappa_{12}^2,\cdots)e^{-\Sigma\alpha\kappa^2}, \quad (4.25)$$

where the *spectral function* σ is given by the finite sum

$$\sigma_{m_1\cdots m_N}(\kappa_{12}^2\cdots) = \sum_{n_{12}\cdots} \prod_{\text{lines}} \frac{\theta(\kappa^2-m^2n_{ij}^2)\Omega_{n_{ij}}(\kappa_{ij}^2)}{n_{ij}!} \times \prod_{\text{vertices}} v(m_k+\sum_{l\neq k} n_{kl}). \quad (4.26)$$

It must be emphasized, however, that the spectral integral (4.25) converges only for *sufficiently large* α_{ij} where it defines an analytic function. If the α's are decreased until a singularity of the function $\mathbf{F}$ is reached, then (4.25) will of course diverge. This happens for the class of theories considered in this paper because the spectral function defined by the sum (4.26) tends to increase like $\exp\kappa^2$.

Although the integral representation (4.25) is not valid for the entire range of the α's, it provides a very useful tool for the analysis of singularities in the momentum-space amplitudes. This is because, according to the integral representation (4.13), only the behavior of $\mathbf{F}(\alpha)$ for *large* α is relevant to the finite p-space structure. Thus if the α-space integrals are divided into two pieces $0<\alpha_{ij}<R$ and $R<\alpha_{ij}<\infty$, then the former yields an *entire* function of the momentum variables while the latter yields the expression

$$\begin{aligned}\tilde{F}_{m_1\cdots m_N}{}^{R}(p_1\cdots p_N) &= \int_R^\infty \prod (d\alpha_{ij})\mathbf{F}_{m_1\cdots m_N}(\alpha_{12}\cdots)\mathbf{C}(\alpha)^{-2}\exp\left(\frac{\mathbf{D}(\alpha,p)}{\mathbf{C}(\alpha)}\right) \\ &= \int_0^\infty \prod (d\kappa_{ij}^2)\sigma_{m_1\cdots m_N}(\kappa_{12}^2\cdots) \\ &\quad\times \int_R^\infty \prod (d\alpha_{ij})\mathbf{C}(\alpha)^{-2}\exp\left(\frac{\mathbf{D}(\alpha,p)}{\mathbf{C}(\alpha)} - \sum_{i<j}\alpha_{ij}\kappa_{ij}^2\right).\end{aligned} \quad (4.27)$$

The integral over α_{ij} contained here approximates to the simple Feynman amplitude corresponding to a diagram with N vertices joined pairwise by lines which correspond to the propagation of bare particles with mass κ_{ij}^2. It certainly has all the usual singularity structure that is proved for perturbation amplitudes. Since the amplitude (4.27) is just a summation of these simple processes weighted by a spectral function which is itself given by a sum of simple phase-space integrals, it seems at least plausible that the requirements of unitarity are met.

It must be remarked, however, that the formal interchange of κ^2 and α integrations employed in arriving at the result (4.27) is not usually permissible. The κ^2 integrals as written are divergent. This difficulty can be met simply by cutting off these integrals at some large mass M^2. Such a cutoff will not affect the singularity structure in any given range of the external momenta if M^2 is taken sufficiently large. One of course takes the limit $M^2\to\infty$ in the end.

PHYSICAL REVIEW D VOLUME 9, NUMBER 4 15 FEBRUARY 1974

Comment on the computation of effective potentials

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England

J. Strathdee
International Centre for Theoretical Physics, Trieste, Italy

(Received 31 May 1973)

The recent result of Coleman and Weinberg deriving the one-loop quantum-corrected effective potential in a self-coupled scalar field theory is rederived using the Dyson-Schwinger equation.

In a recent article Coleman and Weinberg[1] made the interesting point that the spontaneous violation of a Lagrangian symmetry could well be a purely quantum effect. They showed this by computing the one-loop corrections to the equations of motion in a simple model with quartic self-couplings of a set of massless scalar fields. The vehicle for this proof was an effective potential which they computed by summing all connected and one-particle irreducible one-loop graphs with *zero-momentum external lines*. It turns out that this quantum-corrected potential can have a qualitatively different character from the purely classical component. For the case they consider, the new potential favors the emergence of a symmetry-breaking ground state in situations where the purely classical part does not.

The purpose of this note is to remark that the computation of effective potentials and of effective actions in general may be profitably undertaken within the formal framework of Dyson-Schwinger equations. This remark will come as no surprise to those who are familiar with functional methods, but it seems to us that the simplicity and formal power of the method deserves wider recognition.[2] The amusing thing is that, in the approximation studied by Coleman and Weinberg, the functional differential equation which defines the effective action reduces to an *ordinary* (albeit nonlinear) differential equation. Although it may be exceedingly difficult to find exact solutions to this equation, it is certainly possible to develop a solution in powers of $\hbar$. The terms of this semiclassical development are obtained by simple quadratures.

Given a system of fields $\phi^i(x)$ whose dynamics is governed at the classical level by an action functional $S(\phi)$, the basic problem is to construct an *effective action* functional $W(\phi)$ which includes the quantum corrections. This effective action can be obtained, in principle, by solving the Dyson-Schwinger functional differential equations. Here we sketch very briefly the derivation of these equations.

Firstly, if the classical system is perturbed by the introduction of an external source term

$$-\int dx\, J_i(x)\phi^i(x)$$

into the action, the corresponding vacuum amplitude, $\exp[(i/\hbar)Z(J)]$, is given, for example, by the path integral

$$\exp[(i/\hbar)Z(J)] = \int (d\phi)\exp\left\{(i/\hbar)\left[S(\phi) - \int dx\, J_i(x)\phi^i(x)\right]\right\}. \tag{1}$$

The details of how this integral is to be defined are not important. One need suppose only that integration by parts is permissible, viz.,

$$0 = \int (d\phi)\frac{\delta}{\delta\phi(x')}\exp\left\{(i/\hbar)\left[S(\phi) - \int dx\, J_i(\phi)\phi^i(x)\right]\right\}.$$

It is then straightforward to deduce the functional differential equation[3]

$$S_{,i}(\phi)\big|_{\phi \to -\delta Z/\delta J(x) - (\hbar/i)\delta/\delta J(x)} = J_i(x), \tag{2}$$

where the operator $\delta/\delta J$ acts always to the right, the whole expression operating on the identity functional; the variational derivative $\delta S(\phi)/\delta\phi^i(x)$ is abbreviated $S_{,i}(\phi)$. This equation can presumably be given a rigorous meaning if $S(\phi)$ is sufficiently regularized.

The solution of (2), subject to appropriate boundary conditions, should in principle yield the connected vacuum-to-vacuum transition amplitude $(i/\hbar)Z(J)$. An alternative version of this equation, in which the *dependent* variable $Z(J)$ is replaced

by its Legendre transform, the effective action

$$W(\phi) = Z(J) + \int dx\, J_i(x)\phi^i(x), \tag{3}$$

is more convenient, however. The new *independent* variable $\phi^i(x)$ is defined by

$$\phi^i(x) = -\frac{\delta Z(J)}{\delta J_i(x)}. \tag{4}$$

The inverse mapping is clearly given by

$$J_i(x) = \frac{\delta W(\phi)}{\delta \phi^i(x)}. \tag{5}$$

The connected vacuum amplitude $Z(J)$ is of course the generating functional for connected Green's functions. The effective action, $W(\phi)$, on the other hand, is the generating functional for one-particle irreducible vertices (see, for example, Jona-Lasinio[2]). In particular, since the Jacobian matrix $\delta\phi^i(x)/\delta J_j(x')$ is evidently the inverse of $\delta J_j(x')/\delta\phi^i(x)$, it follows that the inverse of the propagator,

$$G^{ij}(x,x'|J) = \frac{\delta^2 Z(J)}{\delta J_i(x)\delta J_j(x')}, \tag{6}$$

is given by

$$G_{ij}{}^{-1}(x,x'|\phi) = -\frac{\delta^2 W(\phi)}{\delta\phi^i(x)\delta\phi^j(x')}. \tag{7}$$

In passing from J_i to ϕ^i, as independent variable we can use the identity

$$\frac{\delta}{\delta J_i(x)} = -\int dx' G^{ij}(x,x'|\phi)\frac{\delta}{\delta\phi^j(x')}, \tag{8}$$

which follows from (4) and (6). It is necessary now to regard G^{ij} as a functional of ϕ rather than J. That is to say, we shall regard G^{ij} as being *defined by the inverse relation* (7).

The Dyson-Schwinger equation for $W(\phi)$ is now simply obtained from (2) by substituting from (4), (5), and (8):

$$\frac{\delta W(\phi)}{\delta\phi^i(x)} = S_{,i}\left[\phi^j(x) + \frac{\hbar}{i}\int dx' G^{jk}(x,x'|\phi)\frac{\delta}{\delta\phi^k(x)}\right]. \tag{9}$$

This equation, together with the definition (7), is to be regarded as the basic functional differential equation for the effective action $W(\phi)$.

Corresponding to the simple case of a single massive scalar field with quartic self-interactions,

$$S(\phi) = \int dx\left[\tfrac{1}{2}(\partial_\mu\phi)^2 - \tfrac{1}{2}m^2\phi^2 - (\lambda/4!)\phi^4\right], \tag{10}$$

the Dyson-Schwinger equation takes the form

$$\frac{\delta W(\phi)}{\delta\phi(x)} = \frac{\delta S(\phi)}{\delta\phi(x)} - \frac{\hbar}{i}\frac{\lambda}{2}\phi(x)G(x,x|\phi) - \left(\frac{\hbar}{i}\right)^2\frac{\lambda}{6}\int dy_1 dy_2 dy_3 G(x,y_1|\phi)G(x,y_2|\phi) \times G(x,y_3|\phi)\frac{\delta^3 W}{\delta\phi(y_1)\delta\phi(y_2)\delta\phi(y_3)}. \tag{11}$$

This equation can be solved by iteration in $\hbar$. The coefficient of $\hbar^n$ in the solution would be represented by the set of n-loop irreducible vacuum graphs, in which the lines correspond to the classical propagator $G_0(x,x'|\phi)$ defined by

$$G_0{}^{-1}(x,x'|\phi) = [\partial^2 + m^2 + \tfrac{1}{2}\lambda\phi(x)^2]\delta(x-x'),$$

and the three- and four-leg vertices to the third- and fourth-order variational derivatives of S,

$$\Gamma_3(x_1,x_2,x_3|\phi) = -\lambda\phi(x_1)\delta(x_1-x_2)\delta(x_2-x_3),$$

$$\Gamma_4(x_1,x_2,x_3,x_4) = -\lambda\delta(x_1-x_2)\delta(x_2-x_3)\delta(x_3-x_4).$$

Examples of such semiclassical developments of the effective action have been treated in great detail by DeWitt.[2]

A method for dealing with (9), which is not semiclassical but which may, perhaps, be appropriate to the treatment of low-energy phenomena, is the following proposal. The zeroth-order approximation here is assumed to be a local functional of $\phi(x)$ given by

$$W_0(\phi) = S(\phi) - \int dx\, V(\phi(x)), \tag{12}$$

where V depends on $\phi(x)$ but *not its derivatives*. If this expression is substituted into (9) one obtains, *for constant* ϕ, the equation

$$\frac{dV}{d\phi} = \frac{\hbar}{i}\frac{\lambda}{2}\phi G(xx|\phi) - \left(\frac{\hbar}{i}\right)^2\frac{\lambda}{6}\left(\lambda\phi + \frac{d^3V}{d\phi^3}\right) \times \int dy\, G(xy|\phi)^3, \tag{13}$$

where $G(xy|\phi) = G(x-y|\phi)$, for constant ϕ, is defined by

$$G^{-1}(x-y|\phi) = \left(\partial^2 + m^2 + \frac{\lambda}{2}\phi^2 + \frac{d^2V}{d\phi^2}\right)\delta(x-y) \tag{14}$$

and is simply the free-field propagator with the effective mass

$$\left(m^2 + \frac{\lambda}{2}\phi^2 + \frac{d^2V}{d\phi^2}\right)^{1/2}.$$

The x dependence in (13) is spurious; since ϕ is a constant we can use translation invariance to set $x = 0$. Thus, it appears that $V(\phi)$ is to be obtained by solving a *second-order ordinary differential equation* for $dV/d\phi$.

The equation is highly nonlinear and no doubt extremely difficult to solve exactly. Moreover, since the local approximation (12) is probably quite unrealistic for the many-loop contributions to W, such exact solutions may not have great relevance. Perhaps a better starting approximation could be made by allowing V to depend on $\partial_\mu\phi$ as well.

If this proposal is to have any value it must be incorporated as the first step in a well-defined approximation scheme. This we have certainly not done, and the above remarks should therefore be taken as nothing more than a preliminary suggestion of a way to go beyond the standard expansion in powers of Planck's constant.

The procedure of Coleman and Weinberg is perturbative in the strict sense in that all graphs of a given order in $\hbar$ which contribute to some part of the effective action are taken into account. However, in the one-loop approximation at least, it is easy to verify that the solution of our nonlinear equation for the effective potential coincides with their result. (To this extent the equation is therefore reliable.) The one-loop contribution to V can be obtained from (13) by a simple quadrature. Thus, neglecting terms of order $\hbar^2$, one finds

$$\begin{aligned} V(\phi) &\approx \int_0^\phi d\phi \frac{\hbar\lambda}{2i}\,\phi G_0(xx|\phi) \\ &= \int_0^\phi d\phi \frac{\hbar\lambda}{2i}\,\phi \int \frac{dk}{(2\pi)^4}\,(-k^2+m^2+\tfrac{1}{2}\lambda\phi^2)^{-1} \\ &= \frac{\hbar}{2i}\int \frac{dk}{(2\pi)^4}\ln\left(\frac{-k^2+m^2+\frac{1}{2}\lambda\phi^2}{-k^2+m^2}\right) \\ &= \frac{\hbar}{2}\int \frac{dK}{(2\pi)^4}\ln\left(\frac{K^2+m^2+\frac{1}{2}\lambda\phi^2}{K^2+m^2}\right), \end{aligned} \tag{15}$$

after making a Wick rotation. This is essentially the result of Coleman and Weinberg. It remains to be seen to what extent this method continues to be reliable in the higher orders.

---

[1] S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973).

[2] The ideas involved are all contained—at least implicitly—in the fundamental papers of J. Schwinger [Proc. Natl. Acad. Sci. U. S. 37, 452 (1951); 37, 455 (1951)], and have been much developed by B. S. DeWitt [*Dynamical Theory of Groups and Fields* (Blackie, London 1965), Chap. 22] and G. Jona-Lasinio [Nuovo Cimento 34, 1790 (1964)].

[3] This derivation is due to D. G. Boulware and L. S. Brown [Phys. Rev. 172, 1628 (1968)].

This page is intentionally left blank

2. Symmetries and Electroweak Unification

Abdus Salam's major independent contribution to particle physics in 1956 was the proposal that a new type of invariance principle, called γ_5-invariance, should hold for the neutrino, thereby predicting that neutrinos should exist only in the left-handed spin state. This so-called two-component theory of the neutrino was also later formulated by Landau and by Lee and Yang. That the neutrino is left-handed is what was found later experimentally. Salam pointed out that parity conservation is incompatible with zero mass of the neutrino and predicted that parity violation in β-decay must be maximal as found experimentally. In an unpublished Imperial College, London, preprint (included in this volume), he extended his γ_5-invariance principle to the four-fermion interaction involving electron and muon, which yielded the correct $V-A$ interaction for leptons but *not* for hadrons. Thus, the γ_5-symmetry principle first formulated by him for neutrinos and other leptons and later extended by Marshak and Sudershan and independently by Feynman and Gell-Mann to the spin-1/2 hadrons in general, resulting in the $V-A$ theory for charged current weak interactions, is now an accepted universal principle for weak interactions.

In the same unpublished paper, in a lengthy footnote, there was an outline of an idea, which one would now call dynamical symmetry breaking in the context of spontaneous breakdown of a symmetry in particle physics, showing how an asymmetrical system can still be described by symmetrical equations. These notions are described succintly nowadays by the statement that the symmetry of the Lagrangian is not a symmetry of the ground state. Later Salam, together with Steven Weinberg and Geoffrey Goldstone, proved Goldstone's conjecture that a massless spin-zero object, called the Goldstone boson, must appear in a theory as a result of the spontaneous breaking of a continuous global symmetry.

Both the γ_5-invariance principle and spontaneous symmetry breaking in the papers referred to above hold in the context of global symmetries. Subsequently, Abdus Salam and John Ward worked on a local gauge theory for the weak and electromagnetic interactions, obtaining the SU(2)×U(1) model in 1964. This was a continuation of their work on the same topic that they had started in 1959. In the intervening period and thereafter, Salam had become deeply convinced that all elementary particle interactions are gauge interactions. This was a recurrent theme in his papers and lectures in the sixties. Some of the papers included in this volume from this epoch are representative of his thinking. Incidentally, Sheldon Glashow in 1961 had also proposed the group SU(2)×U(1) for describing the electromagnetic and weak interactions. However, in the papers of Salam and Ward as well as in Glashow's paper, the problem of generating the masses for the gauge bosons of weak

interactions remained unsolved. So, things remained dormant for some time as far as a viable mechanism for mass generation was concerned.

In the years 1961-64, a lively debate developed about whether the Goldstone theorem could be evaded. P.W. Anderson, using an analogy with the superconductor, pointed out that the Goldstone (plasmon) mode becomes massive due to the gauge field interactions whereas the electromagnetic modes are also massive due to the Meissner effect, despite gauge invariance. However, Anderson had not explicitly given a proof of the evasion of the Goldstone theorem in a relativistic theory. This proof was provided in subsequent theoretical developments pioneered by Peter Higgs, and independently by F. Englert and R. Brout. They also proposed a mechanism by which local symmetries could be broken spontaneously without introducing Goldstone bosons. That this was a way to give masses to the gauge bosons and fermions without introducing explicit mass terms in the Lagrangian was immediately sensed by Salam, and independently by Steven Weinberg. So, all the ingredients mentioned earlier, namely local gauge theory, γ_5-symmetry (leading to chiral fermions), and renormalizability, on which Salam had worked for years, were there. In 1967/68, these developments culminated in the famous papers of Weinberg and Salam resulting in the electroweak unification (a name coined by Salam), based on the SU(2)×U(1) group with spontaneous symmetry breaking. This was a crucial step in the construction of a viable theory of weak interactions, made possible by the imaginative strokes of genius of Higgs, Salam and Weinberg. In 1971, G. 't Hooft proved the renormalizability of the Salam-Weinberg theory, and shortly thereafter B.W. Lee and J. Zinn-Justin gave a proof based on the path-integral formalism, thus removing any residual doubts in the minds of the sceptics.

The SU(2)×U(1) theory with the specific choice of the Higgs fields adopted by Salam and Weinberg made a number of predictions:

- A new kind of neutral current, different from the well known electromagnetic current in quantum electrodynamics (QED), must exist. The Lorentz and (weak) isospin structure of this current were completely specified in the SU(2)×U(1) model.
- In addition to the known spin-1 photon, three additional vector bosons, $W^{\pm}$ and Z^0, mediating, respectively, the charged and the additional neutral weak current, must also exist.
- The seemingly different strengths of the electromagnetic and charged current (Fermi) interactions followed from the large mass splitting between the photon, remaining massless, and the other three massive gauge bosons.
- The $W^{\pm}$ and Z^0 masses were generated by spontaneously breaking the local gauge symmetry from SU(2)×U(1)→$U(1)_{em}$, which was achieved by the introduction of a self-interacting complex scalar field, Φ, transforming as an SU(2)-doublet. The

doublet field Φ and its complex conjugate together comprise four independent fields. Spontaneous symmetry breaking was implemented by giving one of the neutral fields a nonzero vacuum expectation value, $\langle\phi\rangle \equiv \langle 0|\phi|0\rangle = v/\sqrt{2} \neq 0$. Of the four fields in the Lagrangian before spontaneous symmetry breaking, three fields become the longitudinal degrees of freedom of the vector bosons $W^{\pm}$ and Z^0; the photon, coupled to the remaining symmetry group $U(1)_{em}$-generator, remains massless.

- One neutral scalar particle remains in the physical sector of the theory. This is the so-called Salam-Weinberg Higgs particle, which the SU(2)×U(1) model predicts to exist.
- Since the same Higgs doublet is used to give masses to the bosons and fermions, which have Yukawa couplings with the scalar fields, the SU(2)×U(1) model predicts the couplings of the Higgs particle with all the known bosons and fermions but makes no prediction about its mass. This could be traced back to the fact that in the Salam-Weinberg theory, the Higgs particle mass is a function of the unknown quartic Higgs-boson coupling constant.

First evidence for the weak neutral currents came in 1973 from an ISR experiment at CERN using the Gargamelle bubble chamber, in which the first $(\nu_\mu e)$ scattering event was observed. After a brief oscillatory period, the discovery of the weak neutral currents was confirmed by an experiment at Fermilab. The news about the ISR discovery was given to Salam during the European Physical Society meeting held in 1973 at Aix-en-Provence. According to an offen quoted anecdote, Paul Musset of the Gargamelle collaboration, who was driving to the venue of the conference, stopped Salam on the road as he was walking to attend this conference. Apparently, not having met him earlier, Musset enquired if he was Salam. On Salam's affirmative answer, Musset asked him to get into the car as they had discovered the weak neutral current predicted by him and they were going to announce it at the conference!

This was the first triumph of the SU(2)×U(1) model experimentally. Subsequently, experiments measuring the lepton-nucleon and lepton-lepton scattering at CERN, Brookhaven and Fermilab consolidated the evidence in support of neutral currents and established their space-time and isotopic spin properties. Likewise, experiments carried out at reactors using the intense $\bar{\nu}_e$-beams also confirmed the strength and isospin structure of the weak neutral currents. Another classic experiment in this context was carried out in 1978 at SLAC, confirming the observation of parity non-conservation in inelastic scattering of polarized electrons by nucleons, as predicted by the SU(2)×U(1) model. Experimental evidence in support of this model was already convincing and this was recognized through the award of the Nobel prize to Glashow, Salam, and Weinberg for the successful prediction of weak

neutral currents in 1979. The crowning verification of the electroweak model came in 1982 as the massive vector bosons $W^{\pm}$ and Z^0 were discovered in the experiments UA1 and UA2 at CERN, confirming theoretical predictions about their masses and couplings. This indeed was a long shot and can rightly be described as one of the finest examples of theoretical prophecy that this century has witnessed.

The SU(2)×U(1) model for leptons has been extended to the quark sector, with important contributions from Glashow, Iliopoulos and Maiani in 1970 (the GIM mechanism) to cancel the unwanted $O(G_F)$-contribution to the K^0-$\overline{K^0}$ mixing, and by Kobayashi and Maskawa, who extended the electroweak model in 1973 to three families of quarks and leptons, thereby introducing a complex phase in the quark mixing matrix to incorporate CP violation observed in K-decays.

The Standard Model of electroweak interactions, as the Glashow-Salam-Weinberg theory is nowadays called, has been successfully confronted to precise data in the electroweak sector that have come from experiments at LEP and SLC as well as the CERN and Fermilab colliders and νN scattering, carried out laboriously for well over a decade. The agreement between the experiments and the Standard Model, after taking into account the quantum structure of the model, is now completely quantitative, as discussed below. Here, we want to point out that there are still two remaining pieces of this model, namely the top quark and the Higgs boson, that have yet to be discovered. Indirect evidence in favour of the existence of the top quark is compelling, with non-trivial constraints on its mass and couplings from induced effects in the electroweak sector and a number of flavour changing neutral current (FCNC) transitions and direct searches at the Fermilab collider. Likewise, there is no evidence so far speaking against the existence of the Higgs boson. Its mass, however, remains largely undetermined, with LEP experiments providing a lower limit of 63.5 GeV.

In what follows, we review the present status of the quantum (loop) corrected Standard Model in the electroweak sector.

• Present Status of the Standard Model

There exists a wealth of experimental data testing the physics of the Standard Model, ranging in energy from O(eV) (atomic physics) to O(100 GeV) (LEP/SLC and CERN/Fermilab colliders). A detailed discussion of the experiments and the electroweak theory would require several volumes. We shall restrict ourselves here to some of the best tested aspects of the Standard Model using recent reviews on this subject.

The Standard Model has three families of quarks and leptons. As indicated below, they are placed in the left-handed SU(2) doublet and right-handed SU(2) singlet representations of the SU(2)×U(1) group:

$$\text{Leptons: } \begin{pmatrix} \nu_e \\ e \end{pmatrix}_{\rm L}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_{\rm L}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_{\rm L}; \; e_{\rm R}, \mu_{\rm R}, \tau_{\rm R},$$

$$\text{Quarks: } \begin{pmatrix} u \\ d \end{pmatrix}_{\rm L}, \begin{pmatrix} c \\ s \end{pmatrix}_{\rm L}, \begin{pmatrix} t \\ b \end{pmatrix}_{\rm L}; \; u_{\rm R}, d_{\rm R}, c_{\rm R}, \ldots .$$

This quark-lepton replication is needed to cancel the so-called triangle anomalies to make the theory renormalizable. A minimum of three families is needed to incorporate CP violation and accommodate all other aspects of particle physics experiments measured so far. LEP data (and also indirect cosmological limits on N_ν from astrophysical measurements using the primordial He abundance and neutron lifetime) state that this is also very probably the maximum number of quark-lepton families. In particular, experiments at LEP have determined the number of neutrinos to be:

$$N_\nu = 2.98 \pm 0.027 , \tag{7}$$

where a neutrino means any neutral lepton whose mass is less than $m_Z/2$. We recall here that in the Standard Model all three neutrinos are massless though there is no compelling reason for this requirement. While the LEP measurements are very convincing, and all other data are likewise consistent with three neutrino species, it should be stated that a direct measurement of the tau-neutrino, ν_τ, has yet to be made. Experimentally, the following limits hold in the direct searches of nonzero neutrino masses:

$$\begin{aligned} m(\nu_e) &< 7.2 \text{ eV}, \\ m(\nu_\mu) &< 270 \text{ keV}, \\ m(\nu_\tau) &< 31 \text{ MeV}. \end{aligned} \tag{8}$$

In laboratory searches no evidence has been found yet for lepton number violating processes, such as $\mu \to e + \gamma$, and the present upper bounds on the branching ratios translate into (model dependent) lower bounds of $O(1 - 10)$ TeV on the boson masses that could conceivably induce such transitions. While on the topic of neutrino masses and mixings, and hence FCNC transitions in the lepton sector, we remark that the apparent deficit of the solar neutrino flux, compared to the Standard Solar Model expectations in the underground laboratory experiments (the Homestake, Kamiokande, Gallex and SAGE experiments) hints at non-trivial neutrino properties. In particular, the interpretation of the solar neutrino data in terms of nonzero neutrino masses and oscillations in the dense solar matter — the so-called Mikheyev-Smirnov-Wolfenstein (MSW) mechanism — implies lepton number violation. We recall that lepton and baryon numbers are conserved quantum numbers in the Standard Model. So, confirmation of the neutrino mixing would signal physics

beyond the Standard Model. This, however, would not be a radical departure from the SU(2)×U(1) local gauge group, since lepton and baryon number conservation are related to global symmetries and it is easy to break them in, for example, a Grand Unified Theory (GUT). Perhaps, non-perturbative aspects of the Standard Model may also lead to baryon- and lepton-number violations at high energies and temperatures. These non-perturbative effects are, however, not yet quantitative.

In the quark sector, five of the six quarks have been established experimentally. Their electroweak couplings have been measured and found to be in accordance with the Standard Model. Concerning the top quark, indirect evidence in favour of its existence is compelling. The most convincing is the measurement of the weak isospin of the beauty quark, for which present compilations give:

$$I_3(b) = -0.490^{+0.013}_{-0.012}\,. \tag{9}$$

This is entirely compatible with the Standard Model assignment $I_3(b) = -1/2$, implying that the b-quark is indeed in a weak isospin-doublet. Hence, its weak isospin partner, namely the top quark, must exist. Likewise, the top quark plays a dominant role in determining the amplitudes of a number of loop-induced transitions, the most prominent among them being the vacuum-polarization contributions to the $W^{\pm}$ and Z^0 boson masses, and a number of FCNC transitions involving K- and B-decays and mixings as well as CP violation. In fact, determining the top quark mass from electroweak radiative corrections is so far the only useful aspect of these measurements! In the FCNC transitions, also the Cabibbo-Kobayashi-Maskawa (CKM) couplings of the top quark with the three $Q = -1/3$ quarks are measured. All of these couplings are consistent with the unitarity bounds following from the three family structure of the Standard Model. The most stringent lower bound on the top quark mass in direct searches is posted by the CDF collaboration at Fermilab, their limit being $m_t > 113$ GeV at 90 % confidence level. Indirect limits from the precision measurements of the electroweak parameters and quantum effects involving a top-quark loop give:

$$m_t = (164 \pm 27)\ \text{GeV}\,. \tag{10}$$

The other missing particle in the Standard Model is the Salam-Weinberg Higgs boson. There is no model independent theoretical constraint on its mass. This can be readily seen from the Higgs potential of the Standard Model:

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4\,; \quad \mu^2 < 0\,, \quad \lambda > 0\,. \tag{11}$$

Its minimum is at $|\phi| = v/\sqrt{2} = \sqrt{-\mu^2/2\lambda}$, where v is the neutral-Higgs vacuum expectation value. This gives, at the tree level, $m_\phi^2 = 2\lambda v^2$. A bound on the

Higgs mass is equivalent to a bound on the Higgs quartic coupling, which is not determined in the Standard Model. We briefly discuss a number of bounds on m_ϕ, following from well motivated theoretical assumptions.

- **Bounds on the Higgs Boson Mass from Partial Wave Unitarity**

The high energy behaviour of the elastic scattering amplitudes involving the longitudinal $W_{\rm L}, Z^0_{\rm L}$ and/or ϕ bosons must satisfy partial wave unitarity. This requirement can be used to get upper bounds on the Higgs boson mass. Alternatively, it can also be used to set the scale on the centre of mass energy beyond which the unitarity requirement would necessitate physics beyond the Standard Model. We discuss both of these situations below. For the Jth partial wave amplitude, partial wave unitarity implies that $|a_J| \leq 1$ for all J. This bound can be made more restrictive if the partial wave unitarity is used in the form:

$$|a_J|^2 \leq |\mathrm{Im}\, a_J|\,, \tag{12}$$

which in turn implies the relation:

$$(\mathrm{Re}\, a_J)^2 \leq |\mathrm{Im}\, a_J|(1 - |\mathrm{Im}\, a_J|)\,, \tag{13}$$

giving

$$|\mathrm{Re}\, a_J| \leq \frac{1}{2}\,. \tag{14}$$

The best bound emerges for the $J = 0$ partial wave. It was shown some time ago by Lee, Quigg and Thacker that the most stringent unitarity bound is obtained by considering the full coupled channel system involving the $W^+_{\rm L}W^-_{\rm L}, (1/\sqrt{2})Z_{\rm L}Z_{\rm L}, Z_{\rm L}\phi$, and $(1/\sqrt{2})\phi\phi$ scattering amplitude matrix. In the limit $s \gg m^2_\phi \gg m^2_W$ (here s is the centre of mass energy squared in the process indicated), the largest eigenvalue of this matrix then gives the following bound:

$$m^2_\phi \leq \frac{4\pi\sqrt{2}}{3G_{\rm F}} \simeq (0.7\ \mathrm{TeV})^2\,. \tag{15}$$

This is the upper bound on m_ϕ from the use of tree level unitarity in the perturbative approach. It is an interesting and open question how to interpret this bound for m_ϕ obtained in the above stated limit, since it is known that loop effects don't allow an energy independent bound on m_ϕ. It has been argued to constrain the high energy behaviour of the electroweak amplitudes involving the longitudinal $W_{\rm L}, Z^0$ and ϕ bosons in a different limit where $m^2_W, m^2_Z \ll s \ll m^2_\phi$. This is the relevant limit also for the strongly interacting Higgs case (i.e. $\lambda > 1$). Following the argument

given above for the tree level unitarity, the most restrictive bound follows for the (weak) isospin = 0 and $J = 0$ channel, $(1/\sqrt{6})(2W_{\rm L}^{+}W_{\rm L}^{-} \to Z_{\rm L}Z_{\rm L})$. In the above limit this bound can be stated as:

$$s \leq s_c \equiv \frac{4\pi\sqrt{2}}{G_{\rm F}} \leq (1.2\ {\rm TeV})^2\,, \tag{16}$$

where s_c is the maximum s allowed by the tree level unitarity. The correct interpretation of s_c is that it defines the energy scale above which new physics beyond the Standard Model must exist. This partial wave unitarity argument goes beyond the tree level unitarity since the bound derived in (10) can be shown to follow also from non-perturbative low energy theorems. The upper bound on the Higgs mass can be tentatively identified with s_c. This is the scale to be reached in the planned next generation experiments in the parton-parton centre of mass system. The same limit is often described as a *no-lose scenario*, implying that either one would find a Higgs boson much before the partial-wave unitarity limit is reached, or else it would signal breakdown of the perturbation theory in the Standard Model, heralding new physics.

- **Bounds on the Higgs Boson Mass from Triviality Arguments and Lattice Regularization**

Concentrating on the pure scalar part of the Standard Model given by $V(\phi)$ above, and neglecting all other couplings for the time being except the Higgs self-coupling λ in the evaluation of loop effects in $V(\phi)$, one obtains the following renormalization group (RG) based solution for the scalar running coupling constant, $\lambda(Q)$, at the one-loop level:

$$\lambda(Q) = \frac{\lambda(v)}{1 - \frac{3\lambda(v)}{4\pi^2}\ln(Q^2/v^2)}\,. \tag{17}$$

The minus sign in this equation indicates that the Higgs self coupling is not asymptotically free. On the contrary, it implies that regardless of how small $\lambda(v)$ is, $\lambda(Q)$ will eventually blow up at some large energy scale Q (this pole is called the Landau pole). In order to avoid this and to guarantee the positivity of $\lambda(Q)$, one must have

$$\lambda(v) < \frac{2\pi^2}{3}\frac{1}{\ln(Q/v)} \tag{18}$$

giving

$$m_\phi^2 = 2\lambda v^2 = \sqrt{2}\frac{\lambda(v)}{G_{\rm F}} < \frac{2\sqrt{2}\pi^2}{3G_{\rm F}}\frac{1}{\ln(Q/v)}\,. \tag{19}$$

An upper bound on $\lambda(Q)$ implies a corresponding bound on $\lambda(v)$ and hence on m_ϕ. This is the basis of the upper bounds on the Higgs boson mass derived from triviality arguments. In the Standard Model, the running of the coupling constants is a coupled-channel problem where, because of the large top-quark mass, it is important to take into account the running of the top-quark Yukawa coupling. Neglecting the gauge couplings, which only induce nominal quantitative changes, the one-loop RG equations read as follows:

$$\begin{aligned} \frac{d}{dt}h &= \frac{9}{64\pi^2}h^3, \\ \frac{d}{dt}\lambda &= \frac{1}{16\pi^2}(12\lambda^2 + 6\lambda h^2 - 3h^4)\,. \end{aligned} \tag{20}$$

The triviality argument given above for the pure scalar case implies now a bound on the Higgs and top quark masses. A detailed analysis of the coupled-channel case has been carried out, including the next-to-leading order β-functions. Assuming that there is no physics between the electroweak scale and an assumed large mass scale (such as the GUT-scale $O(10^{15})$ GeV or the Planck mass scale $O(10^{19})$ GeV), one gets $m_\phi = O(200)$ GeV. Likewise, non-perturbative effects as the quartic Higgs coupling becomes large have also been estimated, mostly in the context of the lattice-Higgs model. Again, a cutoff on the parameter $\lambda(v)$ provides an upper bound on the Higgs mass. A recent lattice update gives $m_\phi < 640$ GeV, which is actually pretty close to the partial wave perturbative upper bound. Thus, taking into account the present experimental lower bound, a plausible range for the Salam-Weinberg Higgs boson mass is:

$$63.5 \text{ GeV} \leq m_\phi \leq 640 \text{ GeV}\,. \tag{21}$$

While still on the point of the Higgs boson mass, we remark that in the supersymmetric extension of the Standard Model, which is motivated on the stability argument to protect the Higgs boson mass against large quantum corrections from higher scales such as GUTs, in general, there is more than one Higgs field present. Taking into account radiative corrections to the supersymmetric potential, it is argued that the lightest Higgs boson mass in such theories is bounded to lie below $O(150)$ GeV. The discovery of the Higgs particle(s) is clearly the next most important milestone in the understanding of the mass generation puzzle.

It is befitting to end this brief overview of the Standard Model by summarizing the impressive quantitative rapport between the quantum (loop) corrected Standard Model and the data. In general, quantum corrections are renormalized by fixing a specified number of physical quantities at their measured values, which in the present context are usually taken to be the electromagnetic fine structure constant, $\alpha \equiv \alpha_{\rm em}(Q^2 = 0)$, the Fermi coupling constant $G_{\rm F}$, and the mass of the Z^0 boson,

m_Z, being the best of the measured quantities so far. Their present values are:

$$\begin{aligned} \alpha^{-1} &= 137.03598995(61), \\ G_F &= 1.166389(22) \times 10^{-5}\ \text{GeV}^{-2}, \\ m_Z &= 91.187 \pm 0.007\ \text{GeV}. \end{aligned} \tag{22}$$

In addition, in proving the consistency of the Standard Model at the quantum level, a number of other quantities are used as inputs, including the strong interaction (QCD) coupling constant, α_S, and m_W. The W-boson mass, m_W, has also been measured, though at a lesser accuracy than m_Z, with $m_W = 80.22 \pm 0.26$ GeV. The consistency of all the experimental results is then checked in terms of the unknown quantities, namely the top-quark and Higgs-boson masses. It has become customary to express the results in terms of the quantity $\sin^2(\theta_W)$, while specifying the renormalization scheme. We shall not enter here in the technical discussions involving these renormalization schemes but assume that a specific renormalization scheme has been consistently used. In one such relation, the weak mixing angle defined at a given scale, say $\mu = m_Z$, is related to the fine structure constant, α, the Fermi constant, G_F and the Z mass, m_Z, through

$$s^2 c^2 = \frac{A^2}{m_Z^2 (1 - \Delta r)}, \tag{23}$$

where $s^2 \equiv \sin^2 \theta_W(m_Z)$, $c^2 \equiv 1 - s^2$, $A = (\pi\,\alpha/\sqrt{2}\,G_F)^{1/2}$; Δr is the radiative correction factor which is a function of $\alpha, G_F, m_Z, m_t, m_\phi$. The dominant contribution to Δr arises from the light degrees of freedom (quarks and leptons) to the vacuum polarization amplitudes. This contribution is reliably calculable using experimental measurements.

Likewise, taking into account radiative corrections, the relationship between the $W^\pm$ boson mass, $\sin^2 \theta_W$, and the Fermi coupling constant, G_F in the Standard Model also gets modified:

$$m_W^2 = \frac{A^2}{s^2 (1 - \Delta r_W)}, \tag{24}$$

where Δr_W represents another function of the parameters indicated above. The modifications induced in the partial and total widths of the Z^0 have to be taken into account in a similar fashion. The partial decay width $Z^0 \to f\overline{f}$, where f stands for a generic fermion, can be expressed using the so-called improved Born approximation. This can be expressed as

$$\Gamma_f = N_c^f \frac{\hat{\alpha}}{s^2 c^2} \frac{m_Z \rho_f(m_Z^2)}{48} \left[1 + \left(1 - 4|Q_f|k_f s^2\right)^2\right], \tag{25}$$

Table 1. Electroweak observables determined from the combined LEP data and the loop corrected predictions of the Standard Model. The analysis is based on the following bounds: $m_t > 113$ GeV (from CDF), 60 GeV $< m_\phi <$ 1000 GeV, the $m_W(\Delta r_W)$ constraint as discussed in the text, and $\alpha_S(M_Z^2) = 0.123 \pm 0.006$, from the hadronic event topology measured at LEP. The second error in the third column is from α_S. (Source: LEP Electroweak Working Group, D. Schaile et al., 1993.)

Observable	Experimental value	SM prediction
M_Z	91.187 ± 0.007 (GeV)	(input)
$\Gamma(Z \to \text{all})$	2.489 ± 0.007 GeV	$2.488 \pm 0.010 \pm 0.004$ GeV
$\Gamma(Z \to l^+l^-)$	83.82 ± 0.027 MeV	83.7 ± 0.04 MeV
$\Gamma(Z \to \text{invisible})$	497.6 ± 4.3 MeV	500.3 ± 1.9 MeV
$\sigma_0^{\text{hadrons}}$	41.55 ± 0.14 nb	$41.49 \pm 0.03 \pm 0.04$ nb
$R_{\text{hadrons}} = \Gamma_{\text{hadrons}}/\Gamma_{\text{lept}}$	20.76 ± 0.05	$20.78 \pm 0.03 \pm 0.04$
$R_b = \Gamma_b/\Gamma_{\text{hadrons}}$	0.2200 ± 0.0027	0.2160 ± 0.0016
$\sin^2\theta_{\text{lept}}$	0.2321 ± 0.0006	0.2160 ± 0.0013
ρ_{lept}	1.0035 ± 0.0036	1.0026 ± 0.0038

where N_c^f is the colour factor, Q_f the electric charge of the fermion f, and the functions $\rho_f(q^2)$ and $k_f(q^2)$ represent vacuum polarization corrections, with $\hat{e}^2 = 4\pi\hat{\alpha}(m_Z)$ being the electromagnetic coupling at $\mu = m_Z$.

The radiative correction functions $\Delta r, \Delta r_W, (\rho_f(q^2) - 1)$, and $(k_f(q^2) - 1)$, become the basis on which the loop corrected Standard Model and experiments are confronted. These functions involve, among other parameters, the top-quark mass, whose value due to the quadratic dependence of some of these functions on the top-quark mass is numerically important, and the Higgs boson mass, which enters only through logarithms, and hence is not effectively bounded. A combined fit of all the electroweak data has been performed recently by several groups. The result of one of these fits due to Paul Langacker, which is fairly representative, is

$$\sin^2\theta_W(m_Z)_{\overline{MS}} = 0.2328 \pm 0.0007\,. \tag{26}$$

The subscript corresponds to the so-called $\overline{MS}$-scheme of renormalization. The very impressive precision measurements carried out at LEP (and lately also at SLC) allow a more detailed test of the Standard Model. In Table 1, we reproduce the results of a combined fit of all the four experiments at LEP and the (loop) corrected predictions of the Standard Model, recently undertaken by the LEP electroweak working group. The weak angle in this compilation corresponds to the definition:

$$\sin^2\theta_{\text{lept}} \equiv 1 - 4\frac{g_V}{g_A}\,, \tag{27}$$

with g_V and g_A being the leptonic vector and axial vector coupling constants, respectively. The agreement shown in Table 1 is quantitative, with theory and experiment

all lying within one standard deviation. This analysis can be used to put a bound on the top quark mass. A global fit of the LEP data alone in terms of parameters m_t and the QCD coupling constant $\alpha_s(m_Z^2)$, allowing the Higgs boson mass to lie in the range 60 GeV $\leq m_\phi \leq$ 1000 GeV, gives the following numbers:

$$\begin{aligned} m_t &= (166^{+17+19}_{-19-22})\ \text{GeV}, \\ \alpha_S(m_Z^2) &= 0.120 \pm 0.006 \pm 0.002\,. \end{aligned} \tag{28}$$

The same LEP analysis also provides a determination of m_W, giving $m_W = 80.27 \pm 0.13$ GeV, entirely consistent with the direct measurements at the Fermilab collider yielding $M_W = 80.22 \pm 0.26$ GeV. Likewise, the value of $\alpha_S(m_Z^2)$ from the global fit is in very good agreement with the measurements from the hadronic event topology, giving $\alpha_S(m_Z^2) = 0.123 \pm 0.006$. On the other hand, using the LEP data, as well as data from the νN scattering and direct measurements of m_W, give:

$$\begin{aligned} m_t &= (164^{+16+18}_{-17-21})\ \text{GeV}, \\ \alpha_S(m_Z^2) &= 0.120 \pm 0.006 \pm 0.001\,. \end{aligned} \tag{29}$$

The errors on m_t are combined to give the bounds we have already quoted above, $m_t = (164 \pm 27)$ GeV. Ongoing experiments at LEP and SLC will improve the precision on the weak angle (optimistically) by a factor 2. Likewise, experiments at LEP200 would measure m_W with an accuracy of $O(50)$ MeV. The experimental collaborations CDF and D0 at Fermilab are expected to reach the sensitivity on the top-quark mass in the range implied by the LEP electroweak analysis. These anticipated results would confront the Standard Model with tests of unprecedented precision. However, if its present status is any indicator, we should have no doubts that the Standard Model will pass these tests with the same ease and persuasiveness which we are witnessing today. The role of the Standard Model as the Standard Theory of Particle Physics at the electroweak scale is already well established. This still leaves the Higgs particle whose discovery timetable is somewhat hard to predict!

A. SALAM and J. C. POLKINGHORNE
1955, Ottobre
Il Nuovo Cimento
Serie X, Vol. 2, pag. 685-690
© 1955 Società Italiana di Fisica

On the Classification of Fundamental Particles.

A. SALAM

St. John's College - Cambridge, England

J. C. POLKINGHORNE

Trinity College - Cambridge, England (*)

(ricevuto il 18 Giugno 1955)

Summary. — A new four-dimensional isotopic spin formalism is developed for the heavy fundamental particles. All particles correspond to single-valued representations of the rotation group and there is complete symmetry between the isotopic properties of fermions and bosons. The resulting classification is similar to Gell-Mann's except that the degeneracy in the classification of the K-mesons is removed. The forms of the strong interactions are given explicitly. The physical consequences of the theory are discussed in detail. In particular it predicts the existence of a neutral K-meson of lifetime $\sim 10^{-19}$ s. A new experimental technique for detecting this particle is suggested.

1. – Introduction.

PAIS (1) has obtained a classification for the heavy fundamental particles by considering a four-dimensional isotopic spin space. One disadvantage of his scheme is that it predicts the existence of doubly charged hyperons and these have not been observed. In addition, if the reaction: two nucleons → → two hyperons is forbidden (and it has not been observed) there must

(*) Now at the California Institute of Technology, Pasadena, California U.S.A.
(1) A. PAIS: *Proc. Nat. Acad. Sci.*, **40**, 484 (1954).

be two sets of heavy bosons (K-mesons) having the same quantum numbers (2).

A different scheme has been proposed by Gell-Mann (3). However if the experimental evidence that τ-mesons and θ-mesons have different spins and parities is accepted there is again a mysterious degeneracy in the classification of heavy bosons.

The purpose of this note is to develop a new four-dimensional isotopic spin formalism. In the Pais scheme fermions correspond to spinor representations of the rotation group (4) and bosons to tensor representations of the rotation group, in isotopic space. In our scheme all particles correspond to tensor representations and there is complete symmetry in the isotopic properties of fermions and bosons. No multiply charged particles appear and the resulting classification scheme is similar to Gell-Mann's, except that the τ-mesons and θ-mesons are assigned different isotopic spins and so the degeneracy is removed.

In section 2 some mathematical preliminaries concerning the representations of the rotation group in four dimensions are developed. In section 3 this formalism is applied to the problem of classifying the heavy fundamental particles. It is shown how their strong interactions can be written down. In section 4 the physical consequences of the theory for the K-mesons are developed. In particular the theory predicts the existence of a neutral K-meson of lifetime $\sim 10^{-10}$ s. Despite its instability, an experiment to show its existence is suggested.

2. – Isotopic Formalism.

The generators, $I_{\alpha\beta}$ ($\alpha, \beta = 1, ..., 4$), of the rotation group in four dimensions satisfy

$$I_{\alpha\beta} = -I_{\beta\alpha}\,, \tag{1}$$

and

$$[I_{\alpha\beta}, I_{\gamma\delta}] = i\delta_{\alpha\gamma}I_{\beta\delta} + i\delta_{\beta\delta}I_{\alpha\gamma} - i\delta_{\alpha\delta}I_{\beta\gamma} - i\delta_{\beta\gamma}I_{\alpha\delta}\,. \tag{2}$$

If we write

$$\tau_i = \tfrac{1}{2}(I_{4i} + I_{jk})\,, \tag{3}$$

$$\mu_i = \tfrac{1}{2}(I_{4i} - I_{jk})\,, \tag{4}$$

---

(2) M. GELL-MANN and A. PAIS: *Proceedings of the International Physics Conference, Glasgow*, 1954.

(3) M. GELL-MANN: *Phys. Rev.*, **9**, 833 (1954) and reference (2).

(4) Because PAIS takes the charge to be $\tau_3 + \mu_3 + \frac{1}{2}$ (in our notation). This is to get conservation of baryons. However this is guaranteed by their fermion character.

where (i, j, k) is a cyclic permutation of (1, 2, 3), then

$$[\tau_i, \mu_i] = 0 , \tag{5}$$

and

$$[\tau_i, \tau_j] = i\tau_k , \tag{6}$$

$$[\mu_i, \mu_j] = i\mu_k . \tag{7}$$

Equations (6) and (7) are just the commutation relations of the generators of the rotation group in three dimensions. All irreducible representations of these operators may be labelled by a single quantum number, $\tau'(\mu')$, taking integral and half odd integral values. It is well known ([5]) that all irreducible representations of the rotation group in four dimensions can be labelled by the two quantum numbers τ' and μ'. Those representations for which $\tau' + \mu'$ is an integer are the single-valued tensor representations and those for which $\tau' + \mu'$ is half an odd integer are the double-valued spinor representations.

3. – Classification and Interactions.

We write our scheme in terms of the eigenvalues of the operators τ_3 and μ_3: (τ_3 plays the same part as isotopic spin, and μ_3 the same part as « strangeness », in the Gell-Mann scheme.) Electric charge is given by

$$Q = I_{43} = \tau_3 + \mu_3 , \tag{8}$$

for all particles. We classify fermions according to:

	τ_3	μ_3
N^+, N^0	$\frac{1}{2}$, $-\frac{1}{2}$	$\frac{1}{2}$
Ξ^0, Ξ^-	$\frac{1}{2}$, $-\frac{1}{2}$	$-\frac{1}{2}$
Σ^+, Σ^0, Σ^-	1, 0, -1	0
Λ^0	0	0

This is essentially the same as the Gell-Mann scheme. The differences appear in the bosons. We classify according to:

	τ_3	μ_3
π^+, π^0, π^-	1, 0, -1	0
τ^+, τ^0, τ^-	0	1, 0, -1
θ^+, θ^0	$\frac{1}{2}$, $-\frac{1}{2}$	$\frac{1}{2}$
$\bar{\theta}^0$, θ^-	$\frac{1}{2}$, $-\frac{1}{2}$	$-\frac{1}{2}$

([5]) Cf. H. WEYL: *Theory of Classical Groups* (Princeton, 1939).

We postulate that the strong interactions are invariant under rotations in the four-dimensional isotopic space. Since I_{43} and I_{12} commute, they are separately conserved in these interactions and so τ_3 and μ_3 are separately conserved. The interactions with the electromagnetic field are not invariant under rotations but there is no inconsistency in supposing that they conserve τ_3 and μ_3 separately. The weak interactions, causing decays, conserve only $\tau_3+\mu_3$, the total charge.

The nucleons and cascade particles together correspond to the $(\frac{1}{2}, \frac{1}{2})$ representation of the rotation group. Therefore they can be represented by a vector N_α in isotopic space. In the equations of motion we shall have to write a term $a + b\mu_3$ for the mass to give the difference in mass between the nucleon and the cascade particle. The Σ particles correspond to the (1, 0) representation and so can be written as a self-dual antisymmetric tensor $\Sigma_{\alpha\beta}$, and the Λ^0 corresponds to the (0, 0) representation and is a scalar Λ.

Among the bosons the π-mesons are represented by a self-dual antisymmetric tensor $\pi_{\alpha\beta}$, and the θ-mesons are represented by a vector θ_α. The τ-mesons correspond to the (0, 1) representation of the rotation group and so are represented by an anti-self-dual antisymmetric tensor $\tau_{\alpha\beta}$.

In our scheme all particles correspond to single-valued representations of the rotation group. This is essential because only such particles will have integral electric charge. Nearly all representations corresponding to unit charge have experimentally observed counterparts. The only exceptions are a fermion family corresponding to the (0, 1) representation and a scalar boson (0, 0). If the fermion family had a mass greater than about 2800 electron masses it would be highly unstable under θ-decay into a nucleon.

The problem of writing down the strong interactions is simply that of forming appropriate invariants from the vectors and tensors at our disposal. We give two examples, writing only the isotopic indices. The interaction of τ-mesons with nucleons and cascade particles is given by

$$\bar{N}_\alpha \tau_{\alpha\beta} N_\beta \,, \tag{9}$$

and the interaction of θ-mesons with nucleons, cascade particles and Λ-particles is given by

$$\bar{\Lambda}\theta_\alpha N_\alpha + \text{c.c.}\,. \tag{10}$$

4. – K-mesons.

The similarity of our scheme and Gell-Mann's means that many of the predictions are the same. Consequently we shall only discuss in detail the physical properties of the K-mesons since these show up the differences between the two schemes.

We take the τ^0 and its charge conjugate to be same particle since $N^0 \to \Lambda^0 + \tau^0$ is forbidden. However the τ^0 is a highly unstable particle with a fast γ-decay ($\sim 10^{-18}$ s) and a very fast 3π-decay ($\sim 10^{-19}$ s). The mode of production with the lowest threshold is

$$\pi + N \to N + \tau^0 . \tag{11}$$

This should take place above about 0.5 GeV. The existence of this process, together with the other K-meson processes discussed below that take place at higher thresholds, should explain the broad second resonance in high energy $\pi^- - N^+$ scattering ([6]). We note that the reaction (11) corresponds to $\tau' = \frac{1}{2}$. The τ^0 produced will rapidly decay into 3π-mesons and this process may explain the high probability of $N^0 \to \pi^+ + \pi^- + \pi^0$ events in $\pi^- N^+$ scattering at 1·4 GeV ([7]).

A possible way of verifying the existence of the τ^0 despite its short lifetime would be as follows: He atoms are bombarded with π^- with energies just at the threshold for the production of σ_0 by (11). Any τ^0 created will be slow and will certainly decay outside the range of nuclear forces but inside the mesic atomic radius ($\sim 10^{-10}$ cm). A slow π^- produced by this decay would be captured by the remaining proton to form a mesic atom. This way of forming the mesic atom should be much more probable at this energy than direct π^- capture. Consequently there should be a sharp increase in the amount of characteristic mesic atom X-ray emission at the threshold for τ^0 production.

The τ^+ may be produced by reactions such as

$$\pi + N \to \tau^+ + \Xi , \tag{12}$$

while the θ^+ is produced by reactions such as

$$\pi + N \to \theta^+ + \Lambda^0 , \tag{13}$$

$$\pi + N \to \theta^+ + \Sigma . \tag{14}$$

The question whether τ^+ is produced in association with a Ξ or with a Σ or Λ gives another discriminating test between our theory and Gell-Mann's. As yet the evidence ([8]) is too indefinite to decide. The thresholds for (13) and (14)

([6]) M. GELL-MANN and K. M. WATSON: *Annual Review of Nuclear Science*. 4 (1954).

([7]) W. B. FOWLER, R. M. LEA, W. D. SHEPHERD, R. P. SHUTT, A. M. THORNDIKE and W. L. WHITTEMORE: *Phys. Rev.*, 97, 797 (1955).

([8]) W. B. FOWLER, R. P. SHUTT, A. M. THORNDIKE and W. L. WHITTEMORE: *Phys. Rev.*, 98, 121 (1955).

are lower than that for (12) and this agrees with the evidence that θ's are more common than τ's. Since there is some analogy in our scheme between τ's and π's we expect that the K-meson having a weak μ-decay is the τ in analogy with $\pi \to \mu + \nu$.

The τ^- and θ^- can only be produced in pairs with the τ^+ and θ^+ respectively, by such reactions as

$$\pi + N \to \tau^+ + \tau^- + N , \tag{15}$$

$$\pi + N \to \theta^+ + \theta^- + N . \tag{16}$$

This explains the relative abundance of positively charged K-mesons over negatively charged K-mesons, since the former can also be produced by (12), (13) and (14), with lower threshold.

* * *

We wish to thank Mr. W. GILBERT and Mr. R. SHAW for interesting discussions.

RIASSUNTO (*)

Si sviluppa per le particelle pesanti fondamentali un nuovo formalismo assumendo lo spin isotopico quadridimensionale. Tutte le particelle corrispondono a rappresentazioni ad un unico valore del gruppo rotazionale e si ha completa simmetria tra le proprietà isotopiche dei fermioni e dei bosoni. La classificazione risultante è simile a quella di Gell-Mann, salvo che scompare la degenerazione nella classificazione dei mesoni K. Le forme dell'interazione forte sono date esplicitamente. Si discutono dettagliatamente le conseguenze fisiche della teoria. In particolare questa predice l'esistenza di un mesone K neutro di vita media $\sim 10^{-10}$ s. Si suggerisce una nuova tecnica sperimentale per la scoperta di questa particella.

(*) *Traduzione a cura della Redazione.*

ABDUS SALAM
1957, Gennaio
Il Nuovo Cimento
Serie X, Vol. 5, pag. 299-301
© 1957 Società Italiana di Fisica

On Parity Conservation and Neutrino Mass.

ABDUS SALAM
St. John's College - Cambridge

(ricevuto il 15 Novembre 1956)

1. – YANG and LEE[1] have recently suggested that present experimental evidence does not exclude the possibility that parity is not conserved in β-decay. If future experiments confirm this, it may be possible to relate parity-violation in neutrino-decays to the vanishing of neutrino mass and self-mass. The argument is as follows: the free neutrino Lagrangian is invariant for the substitution $\psi_\nu \to \gamma_5\psi_\nu$ ($\bar{\psi}_\nu \to -\bar{\psi}_\nu\gamma_5$). If it is further postulated that neutrino interactions produce no self-mass, one way to secure this is to require that the total Lagrangian also remain invariant for the same substitution [2] (so that $\bar{\psi}_\nu\psi_\nu \to -\bar{\psi}_\nu\psi_\nu$) while other fields (barring degeneracies which we consider later) remain unchanged. In so far as ψ_ν and $\gamma_5\psi_\nu$ have opposite intrinsic parity, most neutrino interactions would then violate parity conservation.

2. – Some consequences of this invariance are noted:

(*a*) To all β-decay couplings [3] $[\bar{p}(x)\Omega n(x)][e(x)\Omega\nu(x)]$ must be added (non-parity conserving) pseudo-couplings $[\bar{p}(x)\Omega n(x)][\bar{e}(x)\Omega\gamma_5\nu(x)]$.

(*b*) The π-decay Lagrangian must read as $f\bar{\mu}(x)i\gamma\partial\pi(x)\ (\gamma_5\nu(x)+\nu(x))+\text{h.c.}$ and similarly for $K_{\mu\nu}$ decay. The fact that neutrino decay interactions violate parity conservation does not depend on whether the processes are direct or take place through other intermediate fields.

(*c*) The magnetic moment of the (Dirac) neutrino must vanish. This is because $\bar{\nu}(x)\sigma_{\alpha\beta}\nu(x) \to -\bar{\nu}(x)\sigma_{\alpha\beta}\nu(x)$. Since $j_\mu = \bar{\nu}(x)\gamma_\mu\nu(x) \to +\bar{\nu}(x)\gamma_\mu\nu(x)$, the neutrino and anti-neutrino are not identical.

---

(1) C. N. YANG and T. D. LEE. The author is indebted for a pre-print.

(2) This is not the mass-reversal transformation discussed by J. TIOMNO (*Nuovo Cimento*, **1**, 226 (1955)) which requires invariance of the total Lagrangian when a substitution $\psi \to \gamma_5\psi$, $m \to -m$ is made for all fields. In this case neutrino self-mass δm_ν need not vanish. In fact $\delta m_\nu \propto mf(m^2)$ and thus $\delta m_\nu\bar{\psi}_\nu\psi_\nu$ does not change sign.

(3) We write all field operators with their particle symbols: thus $\pi(x)$ stands for the π-meson field, $\nu(x)$ for the neutrino-field etc.

(*d*) The conventional μ-decay interactions can be grouped into 2 classes depending on whether two neutrinos or a neutrino and an anti-neutrino are emitted:

$$\text{(A)} \qquad L_A = \sum_{i=1}^{5} g_i [\bar{\mu}(x)\, \Omega e(x)][\bar{\nu}(x)\, \Omega \nu(x)] + \text{h.c.}$$

$$\text{(B)} \qquad L_B = \sum_{i=1}^{5} g_i [\bar{\mu}(x)\, \Omega e^*(x)][\nu(x)\gamma_4\, \Omega \nu(x)] + \text{h.c.}$$

Invariance for the substitution $\nu \to \gamma_5 \nu$ requires that L_A is a sum of vector and axial-vector, and L_B a sum of scalar, tensor and p.s. interactions. For L_B, the tensor interaction can in fact be excluded if we require further that L_B remain unchanged for the interchange of the two neutrino operators. The Michel parameter has, then, the following unique values, irrespective of the strengths of the couplings involved,

$$L_A: \qquad \varrho = \tfrac{3}{4}\,,$$

$$L_B: \qquad \varrho = 0\,.$$

3. – For the π, $K_{\mu\nu}$ and neutron decay interactions it is possible to conserve parity as well as make $\delta m_\nu = 0$, by postulating a parity-degeneracy of electrons and μ-mesons. The substitution operation then reads

$$\nu \to \gamma_5 \nu\,, \qquad \mu_1 \to \mu_2\,, \qquad e_1 \to e_2\,,$$

where the pairs of particles μ_1, μ_2 and e_1, e_2 have opposite intrinsic parities. The fact that parity-degeneracy in electrons has never been observed may be attributed to a local pre-ponderance of electrons of one parity (like the local preponderance of negatively charged electrons.) From the present point of view (i.e. with the demand that parity-conservation hold),

(i) a parity-degeneracy for electrons and μ-mesons makes it unnecessary to require parity-degeneracy for K-mesons. There may or may not be two types of K-mesons and $K_{2\pi}$, $K_{3\pi}$ decays are irrelevant so far as neutrino self-mass is concerned.

(ii) Although it is uneconomical of postulates, one may link a parity-degeneracy of K-mesons with the substitution operation $\nu \to \gamma_5 \nu$, $K_1 \to K_2$, $e_1 \to e_2$, $\mu_1 \to \mu_2$. This in turn must lead to two types of Λ's and Σ's if $\delta m_\nu = 0$. The physically observed $K_1 \pi\pi$ and $K_2 \pi\pi\pi$ decays, however, would now give a non-zero self-mas for the neutrino, in so far as they are not substitution invariant.

4. – The alternative is to give up the idea of parity-conservation. There are now two distinct ways in which parity violation occurs, (*a*) through the requirement that neutrino self-mass vanish; (*b*) in $K\pi\pi$ and $K\pi\pi\pi$ decays. These decays are unrelated to (*a*) and do not affect it.

This is unsatisfactory. The ideal would have been to link all parity violations with the « neutrino-gauge » requirement $\delta m_\nu = 0$. The only suggestion that the author can make is the following:

Assume from strong interactions that a parity degeneracy exists for K-mesons

(and for Λ and Σ hyperons) and that the global (strong as well as weak) Lagrangian is invariant for the substitutions $\nu \to \gamma_5 \nu$, $K_1 \to K_2$, $\Lambda_1 \to \Lambda_2$, $\Sigma_1 \to \Sigma_2$. Parity violation take places for weak-decays in a specified manner which makes the neutrino self-mass (like the photon self-mass) vanish, while decay times for K_1, K_2 (or Λ_1, Λ_2, or Σ_1, Σ_2) particles are all the same.

* * *

The author is deeply indebted to Professor PEIERLS who first suggested investigating the consequences of the requirement $\delta m_\nu = 0$. He has profited from discussions with Dr. PURSEY.

Reprinted from Imperial College preprint.

On Fermi Interactions

ABDUS SALAM
Imperial College of Science and Technology, London

1. With recent experiments on parity-violation[1] and their theoretical interpretation,[2] it appears that some of the difficulties in the hypothesis of a Universal Fermi interaction between (pn), $(\nu\mu)$ and (νe) pairs can be resolved. In this note a specific form of the hypothesis based on recent μ-decay[1] experimental results is presented.

Let $\Omega^1 = 1$; $\Omega^2 = \gamma_\mu$;

$$\Omega^3 = i\gamma_2\gamma_3,\ i\gamma_3\gamma_1,\ i\gamma_1\gamma_2,\ i\gamma_1\gamma_4,\ i\gamma_2\gamma_4,\ i\gamma_3\gamma_4;\ \Omega^4 = i\gamma_5\gamma_\mu\,;$$

$\Omega^5 = \gamma_5$ be the 16 hermitian Dirac matrices. Write $L^k(1, 2; 3, 4) = (\bar\psi^1\Omega^k\psi^2)(\bar\psi^3\Omega^k\psi^4)$ where ψ^1, ψ^2, ψ^3, ψ^4 are four independent Fermi fields. It is well known[3] that

$$L^i(1, 4; 3, 2) = a^{ij}L^j(1, 2; 3, 4) \tag{1}$$

where

$$-4a^{ij} = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & -2 & 0 & 2 & -4 \\ 6 & 0 & -2 & 0 & 6 \\ 4 & 2 & 0 & -2 & -4 \\ 1 & -1 & 1 & -1 & 1 \end{vmatrix}.$$

The result is unaltered if either of $\bar\psi^1$ or $\bar\psi^3$ is replaced by $\bar\psi^1\gamma_5$, $\bar\psi^3\gamma_5$.

In the sequel, we are concerned with paired operators $(\bar p\Omega^k n)$, $(\bar\psi(1-\gamma_5)\Omega^k\mu^+)$, $(\bar\psi(1-\gamma_5)\Omega^k e^+)$.[4] p, n, μ, e stand for the field corresponding to the particle con-

[1] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Phys. Rev. (in the press). R. L. Garwin, L. Lederman and M. Weinrich, Phys. Rev. (in the press). V. Telegdi, Phys. Rev. (in the press).

[2] Abdus Salam, Il Nuovo Cim. V, 229 (1957). T. D. Lee and C. N. Yang, Phys. Rev. (in the press). L. D. Landau, Nuc. Phys. (in the press).

[3] See H. Umezawa, Quantum Field Theory, North Holland Publishing Co., Amsterdam (1956), p. 119. The additional negative sign on the left-hand side arises if one assumes (as here) that all Fermi Fields anti-commute. Ω^i, $i = 1, \ldots, 5$ are respectively the S, V, T, A and P interaction matrices.

[4] The ordering of the operators is based on the assumption that consistently $n \to p + e^- + \bar\nu$, $\mu^+ + n \to p + \bar\nu$ $(\pi^+ \to \mu^+ + \nu)$, and $\mu^+ \to e^+ + \nu + \bar\nu$. If it turns out that $\pi^+ \to \mu^+ + \bar\nu$, this decay could not be comprehended as due to the universal Fermi interaction. With a change of notation the results of this paper would hold if consistently $n \to p + e^- + \nu$, $\mu + n \to p + \nu$, $\mu \to e + \nu + \bar\nu$.

cerned; ψ satisfies $\gamma(\partial/\partial x)\psi = 0$ and $(1+\gamma_5)\psi$ is the anti-neutrino-field in the 2-component theory. $\left[\gamma_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix},\ \gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\right]$ Define $L^k(p, n; \nu, \mu) = (\bar{p}\Omega^k n)(\bar{\psi}(1-\gamma_5)\Omega^k \mu)$; $L^k(\mu, \nu; \nu, e) = (\bar{\mu}\Omega^k(1+\gamma_5)\psi)(\bar{\psi}(1-\gamma_5)\Omega^k e)$; $L^k(p, n; \nu, e) = (\bar{p}\Omega^k n)(\bar{\psi}(1-\gamma_5)\Omega^k e)$. The universal interaction hypothesis will be stated in terms of L^k.

2. In its simplest form the μ-decay interaction can be written as:

$$L^1 = [g_V \bar{\mu}\gamma_\mu e + g_A \bar{\mu}\gamma_5\gamma_\mu e][\bar{\psi}\gamma_\mu(1+\gamma_5)\psi]. \tag{2}$$

The crucial quantitative result of Lederman's experiment is that $g_A \neq g_V$. If g_A strictly equals g_V, we notice that L^1 is invariant for the substitution $e^+ \to -\gamma_5 e^+$ (or $\mu^+ \to -\gamma_5\mu^+$). In so far as the electromagnetic interaction $(\bar{e}\gamma Ae)$ as well as the free electron Lagrangian remain unchanged if

$$e \to -\gamma_5 e, \quad m_e \to -m_e,^5 \tag{3}$$

it would appear that invariance for this substitution is a universal law for electrons and possibly also for μ-mesons.

Conversely, we may postulate that all electrons (and μ-meson) interaction should possess this symmetry. One may then deduce from dimensions that

$$g_V = a + \frac{m_e m_\mu}{M^2} b, \quad g_A = a - \frac{m_e m_\mu}{M^2} b. \tag{4}$$

Here M is a mass. Lederman's experiment shows that (for $a \approx b$) $M > m_e,\ m_\mu$.

At least for the electron, the substitution $e \to -\gamma_5 e$ finds a deeper basis, if it is assumed that the entire physical mass of the electron is electromagnetic in origin (i.e. the bare mass is zero). The Lagrangian then contains no mass term and the bare coupling constants g_A^0 and g_V^0 must be equal. From this point of view the mass term in the free Lagrangian (as well as the 'correction' term $(m_e m_\mu/M^2)b$ in (4)) arise from a cut-off which must be imposed on the theory.[6]

---

[5] This is the mass-reversal transformation studied by J. Tiomno (Il Nuovo Cim. 1, 226 (1955).)

[6] To illustrate how these quantities may be calculated consider only the electromagnetic interaction. If e_0 is the unrenormalized electron field, the Lagrangian is $\bar{e}_0\gamma\partial e_0 + ie_0\bar{e}_0\gamma A_0 e_0$. This is invariant for $e_0 \to -\gamma_5 e_0$. In terms of renormalized quantities rewrite it as:

$$\bar{e}(\gamma\partial + m)e - m\bar{e}e + ie\bar{e}\gamma Ae.$$

If a cut-off Λ^2 is introduced, Λ^2/m^2 satisfies the equation $1 = \int_0^{\Lambda^2/m^2} \rho_2(x^2)dx^2 / \int_0^{\Lambda^2/m^2} x\rho_1(x^2)dx^2$ (H. Lehmann, Il Nuovo Cim. 11, 342 (1954).). In perturbation theory $1 = (3\alpha/4\pi)\log\Lambda^2/m^2$. Thus with a universal Λ^2, the scale of electromagnetic masses is fixed.

The possibility of $m_e \to 0$ (and possibly $m_\mu \to 0$) (if there is no electromagnetic field) justifies non-inclusion of terms containing $M^2/m_e m_\mu$, M_μ/m_e, m_e/m_μ in (4).

Using invariance principle (3), the β-decay interactions fall into two distinct classes; (S, T, P) and (A, V). The general interaction reads

$$L^2 = \left[\sum_{k=1,3,5} g^k L^k(p, n; \nu, 0)\right] + (m_e/M)\left[\sum_{k=2,4} g^k L^k\right]^7 . \tag{5}$$

Note the dimensional factor m_e/M in front of the second bracket. The same applies to $(pn)(\nu\mu)$ interaction:

$$L^3 = \left[\sum_{k=1,3,5} g'^k L^k(p, n; \nu\mu)\right] + (m_\mu/M)\left[\sum_{k=2,4} g'^k L^k\right] . \tag{6}$$

Similarly in the present notation, μ-decay reads[8]:

$$L^1 = \left[\sum_{k=1,2,3} g''^k L^k(\mu, \nu; \nu, e)\right] + m_e m_\mu/M^2\left[\sum_{k=2,4} g''^k L^k\right] . \tag{7}$$

The connection between (7) and (2) is obtained using (1). Noting $L^1(\mu,\nu;\nu,e) = -L^5$, $L^2 = -L^4$, one can verify that

$$g''^5 - g''^1 = 2a,\ g''^2 - g''^4 = b . \tag{8}$$

From Lederman's work, $g''^1 \neq g''^5$.

We may now state the universal Fermi-interaction hypothesis in the following form:

1. $g^i = g'^i = g''^i,\quad i = 1,3;\ g^5 = g'^5 = -g''^5$
2. If all electron and μ-meson interactions are similar, then $g^i = g'^i,\quad i = 2,4.$
3. $g''^1 \neq g''^5$.
4. No statement can be made about $g''^i,\ i = 2,4.$

Some consequences of this hypothesis are:

(i) β-decay proceeds mainly through S, T and P couplings and the Fierz terms are missing.

(ii) In $(pn)(\nu\mu)$ interaction, a relatively large admixture of A and V exists.

---

[7]In the limit $m_e \to 0, e \to -\gamma_5 e$ allows only S, T, P interactions. This is the reason for not including A and V terms in (5) with a factor M/m_e in front.

[8]That S and T interactions in β decay are compatible with $g_A = g_V$ in L^1 was first pointed out by Prof. T. D. Lee. I am grateful to Prof. J. Steinberger for this remark.

3. A π-meson can decay through the $(pn)(e\nu)$ as well as $(pn)(\mu\nu)$ couplings.[9] It is well-known that S, T, V couplings give no decay, P gives a predominant $(e\nu)$ and A a predominant $(\mu\nu)$ decay.

It is indeed possible that $g^5 \equiv 0$ so that no P coupling exists. Even if it turns out from β-decay evidence that $g^5 \equiv 0$, the following reasoning shows why it may not contribute to π-decay. The (Ξ^-, Ξ^0) doublet is similar to the (p, n) doublet, with Ξ^0 presumably the heavier particle[10] and it seems plausible to postulate for it direct Fermi couplings like (6) and (7).

Thus corresponding to the hermitian coupling

$$L^5(p, n; \nu, e) = -g^5(\bar{p}\gamma^5 n - \bar{n}\gamma^5 p)[\bar{\psi}(1-\gamma_5)e + \bar{e}(1+\gamma_5)\psi] \tag{9}$$

there may exist the interaction

$$L^5(\Xi^-, \Xi^0; \nu, e) = -g^5(\bar{\Xi}^-\gamma^5\Xi^0 - \Xi^0\gamma^5\Xi^-)[\bar{\psi}(1-\gamma_5)e + \bar{e}(1+\gamma_5)\psi] \tag{10}$$

with the same coupling constant g^5.

It has been suggested elsewhere[11] that the π interaction with the isotopic spinor (p, n) may be expected to be nearly the same as with (Ξ^0, Ξ^-). Assuming this, we notice that the sign of g^5 is opposite for the two matrix-elements

$$\begin{aligned} \pi^+ &\to p^+\bar{n} \to e + \nu \text{ (or } \mu + \nu) \\ &\to \Xi^0 + \overline{\Xi^-} \to e + \nu \text{ (or } \mu + \nu). \end{aligned} \tag{11}$$

These two matrix-elements may, therefore, very nearly cancel.[12] It is easy to verify that exactly the opposite is true for the axial vector case, with

$$L^4 = -g^4[\bar{p}i\gamma^5\gamma^\mu n + \bar{n}i\gamma^5\gamma^\mu p][\bar{\psi}(1-\gamma_5)i\gamma_\mu e + \bar{e}i\gamma_\mu(1+\gamma_5)\psi], \tag{12}$$

where there is no relative change of sign and the matrix-elements for the π-decay through $(p, \bar{n})$ or $(\Xi^0, \overline{\Xi^-})$ virtual pairs add. For this case, the ratio of decay probabilities is:

$$\frac{\lambda(\pi \to e + \nu)}{\lambda(\pi \to \mu + \nu)} \approx \left(\frac{m_e}{m_\mu}\right)^2 \left[\frac{m_e^2(m_\pi^2 - m_e^2)}{m_\mu^2(m_\pi^2 - m_\mu^2)}\right].$$

[9] See S. B. Treiman and H. W. Wyld, Jr. (Phys. Rev. 101, 1552 (1956)) for a complete discussion of all previous work.

[10] This conjecture is based on arguments similar to those given by Feynman and Speisman for the neutron case (Phys. Rev. 94, 500 (1954)). The decay of $\Xi^0 \to \Xi^- + e^+ + \nu$ (although very rare) would be a very striking event.

[11] B. D'Espagnet, J. Prentki, and Abdus Salam, Nuc. Phys. (in the press).

[12] More precisely, if $f_1\bar{N}N\pi$ and $f_4\bar{\Xi}\Xi\pi$ are the corresponding π-meson interactions with nucleons and cascade particles, the cancellation is exact if $\Sigma(N)/f_1 = \Sigma(\Xi)/f_4$. $\Sigma(N_1)$ and $\Sigma(\Xi)$ are contributions to meson self-energy from nucleon or cascade pairs. This defines an implicit equation for f_4 in terms of f_1. The origin of the cancellation above lies in the possible electromagnetic inversion of the doublet (Ξ^0, Ξ^-) relative to the (p, n) doublet so far as mass is concerned.

The factor in the second bracket arises from the (squared) matrix-elements; the additional factor $(m_e/m_\mu)^2$ comes from the ratio, the coupling constants in L^2 and L^3 and is peculiar to the theory presented here. The ratio of e to μ production in π-decay is therefore $1{:}10^9$.

A quantitative estimate of the absolute π-decay rate using (6) seems to show that $M \sim m_\pi$.[13]

I am indebted to Dr. J. C. Taylor for numerous helpful discussions.

[13] In these estimates one is dealing with diverging integrals and quantitative results are unreliable.

Nuclear Physics **5** (1958) 447—454; © *North Holland Publishing Co., Amsterdam*

ON SYMMETRIES OF WEAK INTERACTIONS

B. D'ESPAGNAT and J. PRENTKI
CERN, *Geneva*
and
ABDUS SALAM
Imperial College of Science and Technology, London

Received 11 October 1957

Abstract: The $|\Delta I| = \frac{1}{2}$ rule is replaced by a conservation law. This new scheme is practically equivalent to the rule just mentioned (with the possibility of introducing in a natural way some $|\Delta I| = \frac{3}{2}$ admixture). Its main advantage is that it can incorporate lepton interactions.

1. Introduction

To give a quantitative description of strange-particle weak decays, several authors have suggested that the weak interaction Lagrangian L_w is an iso-spinor [1–4]). This leads to the selection rule [5]) $|\Delta I| = \frac{1}{2}$, $|\Delta U| = 1$, which has been found fairly successful in various applications †.

† The rule $|\Delta I| = \frac{1}{2}$ is successful in giving a correct account of τ, Λ and Σ^+ decay branching ratios as also of the ratio of Σ^+, Σ^- life-times (if parity violation is taken into account [6])). It provides the only "explanation" of the fact that K^+ has a much longer life than K^0.

The rule is not successful in the prediction of the $(K^0 \to \pi^0+\pi^0)/(K^0 \to \pi^+ + \pi^-)$ branching ratio: the theory predicts a value $\frac{1}{2}$, while the experimental ratio is at best $\approx \frac{1}{4}$. It is, of course, clear that the $|\Delta I| = \frac{1}{2}$ rule is inaccurate to the extent that electromagnetic effects are to be included. It is commonly stated that electromagnetic effects should make only a difference of $\alpha^2 = (1/137)^2$, but this is not necessarily true. Electromagnetic effects appear in the $K^0 \to \pi^0+\pi^0$ and $K^0 \to \pi^+ + \pi^-$ processes through the mass difference of the π^0, $\pi\pm$ mesons. One simple way to see the effect of this mass difference on the matrix-elements for K^0 decay is the following: Let p be the 4-momentum of K^0, $p = p_+ + p_-$ for $K^0 \to \pi^+ + \pi^-$ decay and $p = p_0 + p_0$ for $K^0 \to \pi^0 + \pi^0$ decay. The matrix elements are

$$F_1 = \sqrt{2}\, F(p^2,\ p_+^2,\ p_-^2), \qquad F_2 = F(p^2,\ p_0^2,\ p_0^2).$$

Thus,

$$F_1 = \sqrt{2}\left[F_2 + 2(p_+^2 - p_0^2)\frac{\partial F_2}{\partial p_0^2}\right];$$

assuming from dimensions $\partial F_2/\partial p_0^2 = \text{const.}\ F_2/p_0^2$,

$$F_1 = \sqrt{2}\left[1 + 2 \times \text{const.} \times \frac{p_+^2 - p_0^2}{p_0^2}\right] F_2.$$

If the const. above ≈ 1, we have

$$F_1 \approx \sqrt{2}\left[1 + 4 \cdot \frac{\Delta m}{m_0}\right] F_2 = \sqrt{2} \cdot \frac{25}{22} F_2.$$

Thus, $(F_1/F_2)^2 \approx 625/242$, giving

$$\frac{K \to \pi^0 + \pi^0}{K \to \pi^+ + \pi^-} \approx \frac{3}{8}.$$

Thus, the mass-difference can have a very large effect in this problem.

From a fundamental point of view there are two objections to these suggestions. Firstly, the assumption that L_w is an iso-spinor does not correspond to the conventional procedure in which Lagrangians are normally considered as iso-scalars; secondly leptons do not find any natural place in such a scheme. We would like to suggest an alternative approach to the problem which avoids both these objections, while having the merit of giving practically † the same results as the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. In fact, we shall replace the violation of the rule $|\Delta \mathbf{I}| = \frac{1}{2}$ by a conservation rule [7]).

Briefly, we remark that the global Lagrangian (including all strong and weak interactions) is invariant for the charge-conserving transformations $\phi \to [\exp(iQ\alpha)]\phi$ (Q is the charge). Of the known particles, (π^+, π^0, π^-), $(\Sigma^+, \Sigma^0, \Sigma^-)$; (μ^+, ν, e^-) form charge triplets, while Λ^0 is a charge-singlet. If K and (N, Ξ) interactions can also be written in terms of charge triplets $(K^+, \overline{K}^0 - K^0, K^-)$ $(p, \Xi^0 - n, \Xi^-)$ and charge singlets $(\overline{K}^0 + K^0)$, $(\Xi^0 + n)$, and we shall show this, then Q can be considered as the third component of a vector and charge-conservation is the same thing as rotation invariance for rotations around a given axis in a [3]-space. To the extent that (μ^+, e^-) and (N, Ξ) mass differences can be neglected, we discover that it is possible to consider Lagrangians invariant for full rotations in this [3]-space, so that not only the charge Q (denoted by M_3 in the sequel) but also $\mathbf{M}^2$ is conserved. We find that $\mathbf{M}^2$-conservation possessed by the global Lagrangian when (N, Ξ) and (μ, e) mass-differences are neglected is practically equivalent to the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. In the paper we shall be concerned with the charge-space (**M**-space) and its connection with isobaric-space theories.

The formulation can be given in two different but equivalent forms which correspond to two distinct models for strong interactions, namely the Salam-Polkinghorne [8]) [4]-dimensional theory and the d'Espagnat-Prentki [9]) [3]-dimensional theory. These two models, referred to as S.P. and E.P. theory, will be very briefly reviewed in section 2. It will be emphasized there that a very simple connection exists between them.

2. Theories of Strong Interactions

2.1. THE S.P. THEORY

The isobaric spin space is [4]-Euclidean; only rotations are considered. The different particles are attributed different irreducible representations $\mathscr{D}(I^+, I^-)$ of the rotation group as follows:

$$\begin{array}{lll} (N, \Xi) & \mathscr{D}(\frac{1}{2}\,\frac{1}{2}) & \begin{cases} I_3^- = +\frac{1}{2} \text{ for N} \\ I_3^- = -\frac{1}{2} \text{ for } \Xi \end{cases} \\ \Sigma & \mathscr{D}(1\ 0) & \\ \Lambda & \mathscr{D}(0\ 0) & \end{array} \qquad (1)$$

† *Note added in proof:* Recent experimental data increase the need for some $|\Delta \mathbf{I}| = \frac{3}{2}$ admixture. It is thus satisfactory that such an admixture can be introduced in a natural way in the present scheme (cf. footnote p. 451).

$$\begin{array}{ll} \pi & \mathscr{D}(1\ 0) \\ \mathrm{K} & \mathscr{D}(\tfrac{1}{2}\ \tfrac{1}{2}) \quad \begin{cases} I_3^- = +\tfrac{1}{2} \text{ for K} \\ I_3^- = -\tfrac{1}{2} \text{ for } \overline{\mathrm{K}}. \end{cases} \end{array} \tag{1}$$

Notice that $\mathbf{I}^+$ is just the "ordinary" isobaric spin in three dimensions while $2I_3^-$ is $U = S+N$. (S denotes the strangeness, N the baryon number.) The interactions are classified in the usual way as strong ($\mathbf{I}^+$ and $\mathbf{I}^-$ conserved separately — 4-dimensional invariance) electromagnetic (I_3^+ and I_3^- conserved separately) and weak ($I_3^+ + I_3^- = Q$ conserved). The bare N and Ξ masses are of course assumed to be equal.

2.2. THE E.P. THEORY

The usual 3-dimensional isobaric spin space is considered, rotations and reflexions are taken into account, spinors of 1st and 2nd kinds are defined through $\xi \to i\xi$ $(\mathscr{D}^{(+)}(\frac{1}{2}))$ and $\eta \to -i\eta$ $(\mathscr{D}^{(-)}(\frac{1}{2}))$ in a reflexion through the origin. The attributions are as follows:

$$\begin{array}{lll} \mathrm{N} & \mathscr{D}^{(+)}(\frac{1}{2}) & (\xi) \\ \Xi & \mathscr{D}^{(-)}(\frac{1}{2}) & (\eta) \\ \Sigma & \mathscr{D}^{(-)}(1) & (\text{ps-vector}) \\ \Lambda & \mathscr{D}^{(+)}(0) & (\text{scalar}) \\ \pi & \mathscr{D}^{(-)}(1) & (\text{ps-vector}) \\ \mathrm{K} & \mathscr{D}^{(+)}(\frac{1}{2}) & (\xi). \end{array} \tag{2}$$

If p is the isoparity (reflexion through the origin) then $U = S+N$ is related to p by [10])

$$p = \mathrm{e}^{\frac{1}{2}i\pi U}. \tag{3}$$

Similarly, the reflexion A through the (1,2)-plane, is related to Q by [11])

$$A = \mathrm{e}^{i\pi Q}. \tag{4}$$

Strong interactions are invariant under rotations and reflexions, electromagnetic interactions are invariant under rotations around the 3rd axis and reflexions through the origin (therefore also under reflexions with respect to the (1,2)-plane); weak interactions conserve only $Q = I_3 + \frac{1}{2}U$.

2.3. CONNEXION BETWEEN S.P. AND E.P. THEORIES

A very simple connection exists between these two theories: in the S.P. scheme invariance with respect to rotations around the 3rd axis in $\mathbf{I}^-$-space ensures that the transformation

$$\psi \to \mathrm{e}^{iI_3^-\alpha}\psi \tag{5}$$

carried out for all fields leaves the Lagrangian invariant. For the special value $\alpha = \pi$ this gives, $\mathrm{N} \to i\mathrm{N}$, $\Xi \to -i\Xi$, $\Sigma \to \Sigma$, $\Lambda \to \Lambda$, $\pi \to \pi$, $\mathrm{K} \to i\mathrm{K}$, $\overline{\mathrm{K}} \to -i\overline{\mathrm{K}}$, i.e. just the reflexions of the E.P. theory: this is why $2I_3^-$ and U are in fact the same quantum number.

3. Invariance in Charge-Space

In the S.P. scheme the N and Ξ masses are the same. In order to reproduce the physical mass difference it is essential to break the full 4-dimensional invariance. If one believes in the conservation of ordinary isobaric spin, this should be done by introducing *intermediate* interactions that are invariant under all rotations in $\mathbf{I}^+$-space (and which moreover conserve I_3^-: conservation of strangeness). This is the same as saying that in the E.P. Lagrangian $g_1 \neq g_4$, $g_5 \neq g_7$, . . . This idea has also been considered by Polkinghorne [12]).

As mentioned in section 2, $I_3^+ + I_3^-$ is nothing but the charge Q. Now it is well known that $I_3^+ + I_3^-$ is also J_{12}, one of 6 infinitesimal generators of the 4-dimensional rotation group. Among these, J_{23}, J_{31}, J_{12} form components of a vector with the usual [3] angular momentum commutation rules. Calling $J_{12} = M_3$ etc. we have here another subspace of the rotation group, the [3] **M**-space.

From the fact that Q has only integer values, it follows that the only representations of the [3] rotation group that can have physical sense are the integer ones. In fact, if multiply charged particles are excluded then the only admissible representations are scalar and vector.

Let us recall that our aim is to write down the weak interactions as invariants in a [3]-space. Now the 4-dimensional manifold furnishes three [3] dimensional spaces, namely the $\mathbf{I}^+$, $\mathbf{I}^-$ and **M**-spaces. For obvious reasons both the $\mathbf{I}^+$-space and the $\mathbf{I}^-$-space can be ruled out (as $\Delta I_3^+ = \pm\frac{1}{2}$, $\Delta I_3^- = \pm\frac{1}{2}$, $\Delta U = \pm 1$ for such interactions), thus only the **M**-space remains possible ($\Delta M_3 = 0$ because of charge conservation). *Let us therefore make the assumption that the weak interactions of bare fields are invariant under rotations in* **M***-space.*

Summarizing, the different interactions have the invariance properties given in table 1.

TABLE 1

Interactions of bare fields	Quantum numbers	Rotation-invariance properties
strong (s)	$\mathbf{I}^+$, $\mathbf{I}^-$, I_3^+, I_3^-, **M**, M_3	4-dim.
intermediate (i)	$\mathbf{I}^+$, I_3^+, I_3^-, M_3	3-dim. ($\mathbf{I}^+$ space)
electromagnetic	I_3^+, I_3^-, M_3	
weak (w)	**M**, M_3	3-dim. (**M** space)

As is well known, the $\mathcal{D}(I^+I^-)$ irreducible representation of the [4] rotation group splits into a direct sum of irreducible representations of 3-dimensional rotations through

$$\mathcal{D}(I^+I^-) = \mathcal{D}(I^+ + I^-) \otimes \ldots \otimes \mathcal{D}(|I^+ - I^-|)$$

when only rotations in **M** space are considered. This gives

$$\begin{aligned} (\mathrm{N},\ \Xi) &\to \mathscr{D}(1) \otimes \mathscr{D}(0) \\ \Sigma &\to \mathscr{D}(1) \\ \Lambda &\to \mathscr{D}(0) \\ \pi &\to \mathscr{D}(1) \\ \mathrm{K} &\to \mathscr{D}(1) \otimes \mathscr{D}(0). \end{aligned} \tag{6}$$

More explicitly

$$(\mathrm{N},\ \Xi) \to \begin{cases} \mathbf{B} = \left(\dfrac{\Xi^- + \mathrm{p}}{\sqrt{2}},\ \dfrac{\Xi^- - \mathrm{p}}{i\sqrt{2}},\ \dfrac{\Xi^0 - \mathrm{n}}{\sqrt{2}}\right) \\ B_s = \dfrac{\Xi^0 + \mathrm{n}}{\sqrt{2}} \end{cases} \tag{7}$$

$$(\mathrm{K},\ \overline{\mathrm{K}}) \to \begin{cases} \mathbf{K} = \left(\dfrac{\mathrm{K}^- + \mathrm{K}^+}{\sqrt{2}},\ \dfrac{\mathrm{K}^- - \mathrm{K}^+}{i\sqrt{2}},\ \dfrac{\overline{\mathrm{K}}^0 - \mathrm{K}^0}{\sqrt{2}} \equiv -\mathrm{K}_2{}^0\right) \\ K_s = \dfrac{\overline{\mathrm{K}}^0 + \mathrm{K}^0}{\sqrt{2}} \equiv \mathrm{K}_1{}^0. \end{cases} \tag{8}$$

The other particles $\mathscr{D}(10)$ and $\mathscr{D}(00)$ have of course the same representations in **M** as in $\mathbf{I}^+$ space. It may be worthwhile to point out that the neutral K's which appear here are not the K_0, $\overline{\mathrm{K}}_0$ but the $\mathrm{K}_1{}^0$, $\mathrm{K}_2{}^0$.

Explicit forms of L_w can now easily be obtained. As an example direct weak interactions between baryons and mesons should take the form

$$(\overline{\mathbf{B}} \times \mathbf{\Sigma}) \cdot \boldsymbol{\pi}, \qquad \bar{B}_s\ \mathbf{\Sigma} \cdot \boldsymbol{\pi}; \tag{9}$$

$$\begin{aligned} &(\overline{\mathbf{\Sigma}} \times \mathbf{\Sigma}) \cdot \mathbf{K}, \qquad \overline{\Lambda}\, \mathbf{\Sigma} \cdot \mathbf{K}, \\ &\overline{\mathbf{\Sigma}} \cdot \mathbf{\Sigma}\, K_s, \qquad \overline{\Lambda}\ \Lambda\ K_s. \end{aligned} \tag{10}$$

It is easily verified that these interactions satisfy the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. In fact, the Lagrangian thus obtained is just the one which follows from "spurion" theory with, however, the restriction that the N and Ξ coupling constants are equal †.

When intermediate interactions are introduced to produce the N, Ξ mass difference $\mathbf{M}^2$ is no longer conserved. (The effective N and Ξ coupling

† Strictly speaking, this is only true for interactions containing only one K and no B or conversely. Other **M**-invariant elementary interactions, such as for example the Fermi-like interactions between baryons (e.g. $\overline{\Lambda} B_s\, \overline{\mathbf{B}} \cdot \mathbf{B}$) do not satisfy the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. As far as real processes, such as Λ, Σ or K decays, are concerned, it can be shown that, as a matter of fact, such interactions lead to a violation of the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule only when combined with *intermediate* (as contrasted with *strong*) interactions. On that basis it might be guessed that, even if such weak interactions are present, the violation of the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule would be rather small: it should, however, be stressed that the very possibility of introducing such Fermi interactions in the present scheme is a problem in itself which requires further consideration. It may finally be noticed that weak interactions of a similar kind, namely $(\overline{\mathbf{B}} \times \mathbf{B}) \cdot \boldsymbol{\pi}$, would induce an anomalous Ξ decay ($\Xi \to \mathrm{N} + \pi$).

constants in L_w then become different). Because of the fact, however, that intermediate interactions conserve $\mathbf{I}^+$, the physical L_w still obeys the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. We come back to this question in section 5.

If μ, e, ν mass differences are forgotten, then leptons can easily be introduced into the scheme by making

$$\mathbf{L}\left(\frac{\mu^+ + e^-}{\sqrt{2}}, \frac{-\mu^+ + e^-}{i\sqrt{2}}, \nu\right)$$

the three components of a vector $\mathbf{L}$ in $\mathbf{M}$-space (with the possibility of a second neutrino as a scalar): $\pi \to \mu + \nu$ and $K \to \mu + \nu$ decays can then be accounted for simultaneously. This feature, however, needs some further comment (see section 5).

Weak interaction Lagrangians involving leptons can easily be written; we give here some examples:

$$\begin{gathered}(\bar{\mathbf{B}} \cdot \mathbf{B})(\bar{\mathbf{L}} \cdot \mathbf{L}); \qquad (\bar{\mathbf{B}} \times \mathbf{B}) \cdot (\bar{\mathbf{L}} \times \mathbf{L}) \\ (\bar{\mathbf{B}} \times \boldsymbol{\Sigma}) \cdot (\bar{\mathbf{L}} \times \mathbf{L}) \\ \boldsymbol{\pi} \cdot (\bar{\mathbf{L}} \times \mathbf{L}); \qquad \mathbf{K}(\bar{\mathbf{L}} \times \mathbf{L})\end{gathered} \tag{11}$$

4. Formulation in E.P. theory

The same results can be obtained starting directly from the 3-dimensional space of E.P. theory. The approach is then as follows. The total Lagrangian is

$$L_0 + L_r + L_w$$

where L_0, $L_r = L_s + L_1$ and L_w are the free, the E.P. and the weak interaction Lagrangians respectively: L_0 is invariant under rotations and reflexions in isospace with the E.P. attributions; the same holds for L_r. If, however, the assumption $M_N = M_\Xi$ for bare masses is made, then these are not the only attributions that make L_0 a true isoscalar. The other possibility is to take the integer representations (6) and (7) for $(N\Xi)$ and $(K\bar{K})$, with the supplementary requirement that under reflexions vectorial quantities transform as pseudo-vectors. It is assumed that with this new set of attributions L_w is a true isoscalar †. Then L_w is of course

† We also considered the possibility of taking

$$\left(\Sigma^+, \frac{\Lambda - \Sigma^0}{\sqrt{2}}\right), \left(\frac{\Lambda + \Sigma^0}{\sqrt{2}}, \Sigma^-\right)$$

as isospinors of the 1st and 2nd kind ($M_\Sigma = M_\Lambda$). As shown by Gell-Mann [13]), also the π strong interactions can be made invariant with this choice of representations; the K strong interactions, however, cannot (they should produce $M_\Sigma \neq M_\Lambda$). This choice would make it possible to write weak π interactions as invariants and is therefore a possible alternative to the choice of the present paper. Such a formalism raises, however, a difficulty concerning Σ^- decay (which becomes forbidden in lowest order). These features were considered independently by Ning- Hu [14]). The different behaviours of Σ^+ and Σ^- in this scheme make the choice of ref. [7]) and [15]) and of the present paper, in our opinion, definitely better.

the same as the one in section 3 and therefore leads to the usual $|\Delta \mathbf{I}| = \frac{1}{2}$ selection rule. The point of view chosen here is that a kind of symmetry exists between strong and weak interactions: just as L_0+L_r is a true scalar with the attributions (E.P.) adapted to strong interactions, L_0+L_w is a true scalar with the attributions adapted to weak interactions. Neither of these two sets of attributions makes of course the total Lagrangian invariant under the rotation-reflexion group. There is, however, *one* operation of that group which leaves it invariant: it is the reflexion with respect to the 1, 2 plane. If the operator corresponding to this particular reflexion is A, the general relation between A and the charge operator Q is

$$A = e^{iQ\pi}$$

valid with both sets of attributions. The invariance stated above is thus connected with charge conservation.

5. Discussion

L_w was written in terms of bare fields operators, with the essential assumption $M_N = M_\Xi$ for bare masses. $L_0+L_s+L_w$ is an invariant in **M**: this makes it possible to define a new quantum number $\mathbf{M}^2$ which is a good quantum number as long as intermediate and electromagnetic interactions are neglected. $\mathbf{M}^2$ conservation brings very strong restrictions to L_w, much stronger in fact than the usual $|\Delta \mathbf{I}| = \frac{1}{2}$ rule (it reduces the number of coupling constants by a factor 2). However, intermediate interactions should, of course, not be neglected: they are in fact of paramount importance in making the N and Ξ observed masses different and in giving a meaning to strangeness. As they are not invariant in **M**-space their introduction has the result that the effective L_w no longer has the invariance properties discussed before. The $\mathbf{M}^2$ conservation rule is thus only a poor approximation. As the net influence of intermediate interactions on the N, Ξ masses is approximately 30 %, it can be hoped, however, that the violation of $\mathbf{M}^2$ conservation produced by the introduction of N, Ξ mass differences is also of this order. It should be pointed out that, because of the fact that intermediate interactions conserve $\mathbf{I}^+$, the weak interactions derived from (9), (10) still rigorously obey the $|\Delta \mathbf{I}| = \frac{1}{2}$ rule. In some sense the present scheme is still more restrictive than the $|\Delta \mathbf{I}| = \frac{1}{2}$ selection rule because the coupling constants of the effective L_w could in principle be estimated. It should be stressed, however, that Fermi-like interactions can give a $|\Delta \mathbf{I}| = \frac{3}{2}$ admixture (see footnote p. 451).

The reason for non conservation of $\mathbf{M}^2$ is of course the fact that $\mathbf{I}^+$ does not commute with **M** (nor does $\mathbf{I}^-$) and that, for this reason, a particle cannot be attributed simultaneously I^+ and M values. As, however, the intermediate interactions are much stronger than the weak ones, the I^+,

I_3^- quantum numbers have a much more definite physical significance than the M quantum number. Quite similar remarks apply to the [3] dimensional approach †.

Finally, the present scheme is more powerful than the usual $|\Delta \mathbf{I}| = \frac{1}{2}$ rule in that it can also be applied to leptons. This, however, is only possible if the μ and e bare masses are taken equal. This, in turn, raises a difficulty which is more serious than in the nuclear particles case because no equivalent of the intermediate interactions is here at our disposal.

Of course, if [4]-representations were attributed to leptons this problem might be solved, but the danger is then that strong interactions of leptons might turn up that would contradict experiment. This raises, however, no conceptual difficulty because phenomenological mechanisms can be thought of, which would solve the dilemma: as an example consider the assumption that the only intermediate interaction is $g(\bar{\Xi}^- \Xi^- + \bar{\Xi}^0 \Xi^0)\ (\bar{\mu}\mu)$. This interaction is sufficient to account for the Ξ, N mass difference as well as for the μ, e mass difference. The suitable value for g seems to be rather small in view of the quadratic divergence involved. One would expect an anomalous μ scattering from nucleons whose magnitude is being estimated.

† It needs hardly be pointed out that the philosophy is adhered to that symmetries, invariance properties etc. which may exist in nature are intimately connected with *bare* fields and can be spoiled, or even completely hidden when *physical* particles are considered.

References

1) R. Gatto, Nuovo Cimento **2** (1956) 318
2) B. d'Espagnat and J. Prentki, Nuovo Cimento **3** (1956) 1045
3) G. Takeda, Phys. Rev. **101** (1956) 1547
4) G. Wentzel, Phys. Rev. **101** (1956) 1215
5) R. H. Dalitz, quoted by R. E. Marshak, B.A.S. **30** (1955) 25
6) B. T. Feld, Nuovo Cimento **6** (1957) 650
7) B. d'Espagnat, C. R. Acad. Sc. **245** (1957) 894
8) A. Salam and J. C. Polkinghorne, Nuovo Cimento **2** (1955) 685
9) B. d'Espagnat and J. Prentki, Nuclear Physics **1** (1956) 33
10) G. Racah, Nuclear Physics **1** (1956) 302
11) B. d'Espagnat and J. Prentki, Phys. Rev. **102** (1956) 1684
12) J. C. Polkinghorne, Nuovo Cimento **6** (1957) 864
13) M. Gell-Mann, Phys. Rev. **106** (1957) 1296
14) Ning-Hu, Padova-Venice Conference (1957)
15) B. d'Espagnat, J. Prentki and A. Salam, Padova-Venice Conference (1957)

ABDUS SALAM and J. C. WARD
1959, Febbraio
Il Nuovo Cimento
Serie X, Vol. 11, pag. 568-577
© 1959 Società Italiana di Fisica

Weak and Electromagnetic Interactions.

ABDUS SALAM
Imperial College - London

J. C. WARD
University of Miami - Miami, Flo. (*)

(ricevuto il 3 Dicembre 1958)

Summary. — The postulate of a « local connection » in a [3] charge space leads to the introduction of three spin one fields. One of these can be identified with the electro-magnetic field and the other two can be shown to mediate all known weak interactions, thus unifying these interactions with electro-magnetism. The theory takes account of the fact that weak interactions violate parity and strangeness conservation while electro-magnetic interactions do not do so.

---

1. – D'ESPAGNAT, PRENTKI and SALAM (1) have recently advanced the hypothesis that a three-dimensional [3] charge space Q exists and that all elementary particles correspond to its scalar or vector representations.

These authors show that possibly weak interactions exhibit full rotational symmetry in Q space. Electro-magnetic and strong interactions violate the full symmetry; however these interactions are invariant for rotations around one particular axis in this space (the « charge-axis ») so that Q_z is always conserved.

In this paper we wish to make a fresh approach with the idea of a [3] charge space. Following some ideas first advanced by SCHWINGER (2), we show

---

(*) Visiting Professor at the University of Maryland, present address, Free University, of Brussels.

(1) B. D'ESPAGNAT, J. PRENTKI and A. SALAM: *Nucl. Phys.*, **5**, 447 (1958). This paper will here by referred to as ESP. Similar ideas have been suggested by G. TAKEDA: to be published. In ESP the Q space is designated as M space.

(2) J. SCHWINGER: *Ann. of Phys.*, 2, 407 (1957).

that parity-conserving weak and electromagnetic interactions taken together form a unit which exhibits full rotational symmetry in Q space. This full symmetry is broken on the one hand by parity-violating weak interactions (γ_5-invariance) and on the other hand by strong interactions. We go further in respect of the rotational invariance of electro-magnetic and (parity conserving) weak interactions in Q space; we wish to make the hypothesis that the orientation of all three charge axes can be chosen arbitrarily at all space time points.

Just as it is necessary to introduce the electro magnetic field if it is assumed that the orientation of axes in the conventional [2] ([3]) space spanned by fields Φ and Φ^* (or alternatively by the fields Φ_1 and Φ_2, where $\Phi=\Phi_1+i\Phi_2$) is arbitrary at all space time points, our « three-dimensional gauge invariance » makes it necessary to introduce in Q space a triplet of fields consisting of two charged vector bose fields in addition to the electro-magnetic field ([4]). It will be shown that the form of the interaction of the charged fields is such that they can mediate weak interactions.

2. – Let us first consider the problem of the generalized gauge transformation ([5]). Let ψ represent a vector in charge-space, ($\psi=\psi_1, \psi_2, \psi_3$ where, ψ_1, ψ_2, ψ_3 are 3 Fermi fields), which transforms as

$$\psi' = S\psi \,. \tag{1}$$

If S is a function of x, y, z, t it is necessary for invariance, that all derivatives of ψ appear in the combination

$$(\partial_\mu - i\varepsilon B_\mu)\psi \,, \tag{2}$$

where B_μ are 3×3 matrices, and transform as

$$B'_\mu = S^{-1}B_\mu S + \frac{i}{\varepsilon} S^{-1}\frac{\partial S}{\partial x_\mu} \,.$$

([3]) A priori, it is more attractive to associate charge with a [3] group structure, rather than with a [2] structure as in conventional theory. This is because infinitesimal matrices describing rotations in three dimensions possess three eigenvalues $\pm 1, 0$ (rather than just ± 1), so that the same formulation can give charge values of *all* charged and neutral particles. We are indebted for this remark to Prof. M. FIERZ.

([4]) This was first done by SCHWINGER (ref. ([1])). The difference between our work and Schwinger's is that here the charged Bose fields arise naturally from the 3-dimensional charge-gauge and there is little arbitrariness in writing down the interaction Lagrangian.

([5]) Mathematically the work in this Section in completely analogous with that of YANG and MILLS (*Phys. Rev.*, **96**, 191 (1954)); and R. SHAW (Cambridge dissertation, unpublished (1954)).

YANG and MILLS were considering the possibility of a generalized gauge transformation in three-dimensional isotopic rather than charge-space.

If Q_t are the set of three rotation matrices in Q space, one may write

$$B_\mu = A_\mu(x)\cdot Q$$

and define

$$f_{\mu\nu} = e_{\mu\nu} - S_{\mu\nu}\,, \tag{3}$$

where

$$e_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu}\,, \tag{4}$$

$$S_{\mu\nu} = \varepsilon(A_\mu \times A_\nu)\,. \tag{5}$$

Here $\times$ is the symbol for vector-product.

It is easy to verify that

$$F_{\mu\nu} = f_{\mu\nu}\cdot Q$$

transforms as

$$F'_{\mu\nu} = S^{-1}F_{\mu\nu}S\,. \tag{6}$$

The A field (or alternatively $f_{\mu\nu}$ field) has arisen as a direct consequence of our demand that S depends on x, y, z, t.

A charge-gauge invariant Lagrangian for A field is

$$-\tfrac{1}{4}f_{\mu\nu}f_{\mu\nu}\,. \tag{7}$$

The $S_{\mu\nu}$ terms are necessary for the invariance. Analogously to ref. ([5]), the total lagrangian density is

$$-\tfrac{1}{4}f_{\mu\nu}f_{\mu\nu} - \bar{\psi}\gamma(\gamma - i\varepsilon QA)\psi - m\bar{\psi}\psi\,. \tag{8}$$

Notice that

$$J = \bar{\psi}Q\psi = -i(\bar{\Psi}\times\psi)$$

so that the interaction term in (8) is

$$J\cdot A = \varepsilon\bar{\psi}\chi\gamma_\mu\psi\cdot A\,.$$

It is easy to verify that the subsidiary condition

$$\frac{\partial A_\mu}{\partial x_\mu} = 0\,, \tag{9}$$

is compatible with the equations of motion derived from the Lagrangian; also that the « current »

$$\mathcal{J}_\mu = J_\mu + \varepsilon A_\nu \times f_{\mu\nu} \tag{10}$$

satisfies the equation of continuity

$$\frac{\partial \mathcal{J}_\mu}{\partial x_\mu} = 0 . \tag{11}$$

Thus

$$Q = \int \mathcal{J}_4 \mathrm{d}^3x ,$$

is independent of time.

One can verify that components of A field carry charge $+\varepsilon$, 0, $-\varepsilon$. Picking out an arbitrary charge axis and defining

$$A^\pm = \frac{1}{\sqrt{2}}(A_1 \mp iA_2) , \qquad A^0 = A_3 , \tag{12}$$

and similarly for $\psi^\pm$, ψ^0 we note that

$$\frac{1}{2}(f^1_{\mu\nu}f^1_{\mu\nu} + f^2_{\mu\nu}f^2_{\mu\nu}) = (e^+_{\mu\nu} - S^+_{\mu\nu})(e^-_{\mu\nu} - S^-_{\mu\nu}) =$$

$$= \left[\left(\frac{\partial}{\partial x_\nu} - i\varepsilon A^0_\nu\right) A^+_\mu - \left(\frac{\partial}{\partial x_\mu} - i\varepsilon A^0_\mu\right) A^+_\nu\right]\left[\left(\frac{\partial}{\partial x_\nu} + i\varepsilon A^0_\nu\right) A^-_\mu - \left(\frac{\partial}{\partial x_\mu} + i\varepsilon A^0_\mu\right) A^-_\nu\right],$$

while

$$-\tfrac{1}{4} f^3_{\mu\nu} f^3_{\mu\nu} = -\tfrac{1}{4} e^0_{\mu\nu} e^0_{\mu\nu} + \tfrac{1}{2} S^0_{\mu\nu} e^0_{\mu\nu} - \tfrac{1}{4} S^0_{\mu\nu} S^0_{\mu\nu} .$$

Identifying A^0_μ and $e^0_{\mu\nu}$ with the electro-magnetic field, the lagrangian $-\frac{1}{4} f_{\mu\nu} f_{\mu\nu}$ is a sum of terms which represent

1) the free Maxwell Lagrangian $(-\frac{1}{4} e^0_{\mu\nu} e^0_{\mu\nu})$;

2) the terms: $-\frac{1}{2}(\partial^+_\nu A^+_\mu - \partial^+_\mu A^+_\nu)(\partial^-_\nu A^-_\mu - \partial^-_\mu A^-_\nu)$,

where

$$\partial^\pm = \left(\frac{\partial}{\partial x} \mp i\varepsilon A^0\right).$$

These terms describe the conventional electro-magnetic interaction of charged fields A^+, A^-.

3) Terms $-\frac{1}{4}S^0_{\mu\nu}S^0_{\mu\nu} = \frac{1}{4}\varepsilon^2((A^+_\mu A^-_\nu - A^-_\mu A^+_\nu)^2$ which are in a sense « mass » terms for the charged fields A^+, A^-.

4) A term $\frac{1}{2}S^0_{\mu\nu}e^0_{\mu\nu} = -(i/2)(A^+_\mu A^-_\nu - A^+_\nu A^-_\mu)e^0_{\mu\nu}$.

This term would represent a (Pauli) anomalous magnetic moment for A^+, A^- fields, if these fields had mass.

The A field by itself is then a most remarkable field. One of its components can be identified with the Maxwell field, the other two represent charged particles, carrying « anomalous magnetic moment » while in place of the conventional mass term $-\mu^2 A^+A^-$, there is a more complicated (quartic) term

$$\frac{\varepsilon^2}{2}[(A^+_\mu A^-_\mu)^2 - (A^+_\mu)^2(A^-_\mu)^2] .$$

Even in the presence of interactions with other particles the supplementary condition has the transparent form

$$\frac{\partial A_\mu}{\partial x_\mu} = 0 .$$

The theory is renormalizable; it is perhaps the only theory of charged vector mesons which can be renormalized. Finally note that

$$-\bar{\psi}\times\psi\cdot A = (\bar{\psi}^+\psi^+ - \bar{\psi}^-\psi^-)A^0 + (\bar{\psi}_0\psi^- - \bar{\psi}^+\psi_0)A^+ + \text{h. c.}$$

3. – Let us assume that $L_1 = (e^+, \nu, e^-)$ form a vector in charge-space. With

$$e^\pm = \frac{1}{\sqrt{2}}(\psi_1 \mp i\psi_2) , \qquad \psi_3 = \nu ,$$

(where the three ψ are Majorana fields), the derivative term in the free Lagrangian is

$$\bar{\psi}\gamma\partial\psi ,$$

which leads to an interaction with A field of the type $\bar{\psi}\times\psi\cdot A$. Writing this in full

$$[\bar{e}^+\gamma_\mu\nu - \bar{\nu}\gamma_\mu e^-]A^+_\mu + \text{h. c.} + [\bar{e}^+\gamma_\mu e^+ - \bar{e}^-\gamma_\mu e^-]A^0_\mu . \tag{14}$$

The last term represents electromagnetic interaction while the first two

terms could give the weak interactions of (e, ν) pair. However so far there is nothing in the theory to pick out any one axis in charge-space.

We wish now to do precisely this, by requiring that the neutrino Lagrangian remain invariant for

$$\nu \to \gamma_5 \nu \,. \tag{15}$$

in so far as the neutrino mass must be zero. One result of this is that the first term in L_{int} reads

$$\tfrac{1}{2}[\bar{e}^+\gamma_\mu(1+\gamma_5)\nu - \bar{\nu}\gamma_\mu(1+\gamma_5)e^-]A_\mu^+ \,.$$

Thus the neutrino-gauge, conflicting as it does with the [3] charge-gauge, picks an axis in charge-space, and (incidentally) leads to parity violation of interactions of the A_μ^+, A_μ^- fields. The possibility of gauging the neutrino ($\nu \to \gamma_5\nu$) arises from the fact that neutrino has zero mass while the electron does not. In a sense then the existence of mass is incompatible with full rotational symmetry in [3] charge-space; the fact that all known charged particles have mass is the expression of the fact that one particular axis is preferentially chosen in [3] charge-space.

One can consider a second lepton vector $L_2 = (\mu^+, \nu', \mu^-)$, where ν′ is gauged as $\nu' \to +\gamma_5\nu'$ (*). The interaction Lagrangian for L_2 fields is $-\varepsilon \bar{L}_2 \times L_2 \cdot A$. Through the intermediacy of the A^+, A^- fields, one then has a possibility of $\mu \to e + \nu + \nu'$, all helicities given by our interaction Lagrangian being precisely the ones which are observed.

The major problem that remains is the problem of the mass of $A^\pm$ fields. Before the introduction of the parity violating term in (14), there was nothing in the theory to differentiate $A^\pm$ from A^0. The charged A fields would (in spite of the term

$$-\frac{\varepsilon^2}{2}[(A_\mu^+ A_\mu^-)^2 - (A_\mu^+)^2(A_\mu^-)^2]\Big),$$

acquire no self-mass from its interactions (**). The parity violating terms give the possibility that self-mass for A^+, A^- no longer vanishes. Also the charge-renormalization for A^+, A^- interactions is different than for the A^0 interactions.

(*) In a sense ν and ν′ are twin neutrinos.

(**) This has actually been checked to the lowest order by S. KAMEFUCHI. We are indebted for a private communication. He has also checked that an effective mass will appear for the $A^\pm$ fields if parity-violating terms are added.

We propose to come back to this problem in a subsequent paper. Here we simply note that if $\varepsilon^2/4\pi = 1/137$, A^+, A^- fields would need to have an effective mass $\approx\sqrt{\varepsilon^2/g}$, *i.e.* about $30\,m_p$. Note that a (quadratic) self-mass is $\alpha\varepsilon^2\Lambda^2$, where Λ is the cut-off mass. Thus $\varepsilon^2\Lambda^2 \approx \varepsilon^2/g$ gives $\Lambda \approx 350\,m_p$. This does not appear too excessive a value for the cut-off.

4. – Baryons and mesons.

For π and K mesons, define the charge space vectors $\boldsymbol{\pi}$ and $\boldsymbol{K}$ composed from (π^+, π^0, π^+) and (K^+, K_1^0, K^-) particles. (More precisely $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$,

$$\pi^0 = \pi^3, \quad \pi^\pm = \frac{1}{\sqrt{2}}(\pi^1 \mp i\pi^2), \qquad K_1^0 = \frac{1}{\sqrt{2}}(K^0 + \bar{K}^0).$$

The interaction with the A field is

$$-\varepsilon(\partial\boldsymbol{\pi}\times\boldsymbol{\pi}\cdot\boldsymbol{A}) - \varepsilon^2[\boldsymbol{A}^2\boldsymbol{\pi}^2 - (\boldsymbol{A}\cdot\boldsymbol{\pi})^2]$$

and similarly for the K mesons.

For baryons, we make the fundamental assumption that all bare baryon masses are equal. Defining

$$Y_1 = \frac{1}{\sqrt{2}}(\Lambda^0 + \Sigma^0), \qquad Y_2 = \frac{1}{\sqrt{2}}(\Lambda^0 - \Sigma^0),$$

we write two baryon vectors

$$\boldsymbol{B}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} p + \Sigma^- \\ i(p - \Sigma^-) \\ n + Y_1 \end{pmatrix}, \qquad \boldsymbol{B}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} \Sigma^+ + \Xi^- \\ i(\Sigma^+ - \Xi^-) \\ Y_2 + \Xi^0 \end{pmatrix}.$$

The free baryon Lagrangian can be written as

$$[\bar{\boldsymbol{B}}_1 D\boldsymbol{B}_1 + \bar{\boldsymbol{B}}_2 D\boldsymbol{B}_2 + \bar{B}_{s1} DB_{s1} + \bar{B}_{s2} DB_{s2}] + [\bar{\boldsymbol{B}}_1^c D\boldsymbol{B}_1^c + \bar{\boldsymbol{B}}_2^c D\boldsymbol{B}_2^c + \bar{B}_{s1}^c DB_{s1}^c + \bar{B}_{s2}^c DB_{s2}^c].$$

Here

$$B_{1s} = \frac{1}{\sqrt{2}}[n - Y_1], \qquad B_{2s} = \frac{1}{\sqrt{2}}[Y_2 - \Xi^0],$$

D is the Dirac operator and $\boldsymbol{B}^c$ is the charge-conjugate vector. (Using Majorana representation of Dirac matrices if $\psi = \psi_1 + i\psi_2$ the charge conjugate spinor is $\psi^c = \psi_1 - i\psi_2$).

We note here an essential difference from the case of leptons. Each one

of our vectors $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, etc., contains complex fields as components. For leptons ($\boldsymbol{L}_1$, $\boldsymbol{L}_2$) as well as for π and K mesons, the corresponding charge-space vectors were real. The reason why $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ are complex lies in baryon-conservation. Thus one can adopt two completely different attitudes for baryons; one can consider that there are essentially four independent charge-space vectors $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\boldsymbol{B}_1^c$ and $\boldsymbol{B}_2^c$ and baryon-conservation demands that $\boldsymbol{B}_1$, $\boldsymbol{B}_2^c$ (for example) should occur symmetrically. Alternatively one may split each of the vectors $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ into real and imaginary parts and work with the resulting four vectors. It is crucial that there are four baryons which are charged (P,Σ^-), (Σ^+, Ξ^-). One could never arrive at the ideas of charge-space vectors before the discovery of new particles.

A charge space rotation will produce the interaction Lagrangian.

$$\varepsilon[\overline{\boldsymbol{B}}_1\boldsymbol{B}_1\cdot\boldsymbol{A}+\overline{\boldsymbol{B}}_2\boldsymbol{B}_2\cdot\boldsymbol{A}]-\varepsilon[\overline{\boldsymbol{B}}_1^c\boldsymbol{B}_1^c+\overline{\boldsymbol{B}}_2^c\boldsymbol{B}_2^c]\cdot\boldsymbol{A}\,. \tag{12}$$

The Lagrangian combines formally electro-magnetic interactions, as well as interaction terms which could be responsible (through A^+, A^- fields) for β decays as well as for hyperon decays. However at this stage only parity conserving weak decays are included.

On the other hand we know from our work with leptons, that the postulated γ_5-invariance of neutrino interactions picks out a direction in charge-space and also symmetrizes the weak lepton interactions for parity-conserving and parity-violating terms. It thus appears that neutrino γ_5-invariant Lagrangian is also γ_5-symmetrized in the sense used earlier by SALAM (*).

A term in a Lagrangian is « γ_5-symmetrized » if the trasnformation $\psi\to\gamma_5\psi$, $m\to -m$ are made *independently* for all fields concerned (**) and the result summed. Since the lepton and A field interaction Lagrangian happens to be « γ_5-symmetrized » in the above generalized sense, it seems necessary to « γ_5-symmetrize » the baryon-Lagrangian as well. Thus

$$p,\ Y_1,\ \Sigma^-\to\gamma_5(p,\ Y_1,\ \Sigma)\,. \tag{13}$$

A priori (Σ^+, Y_2, Ξ^-) could transform as $+\gamma_5$ or $-\gamma_5$ but Y_1 and Y_2 are related and thus it appears likely that $(\Sigma^+, Y_2, \Xi^-)\to\gamma_5(\Sigma^+, Y, \Sigma^-)$. After this symmetrization the electro-magnetic interaction remains unchanged, (*e.g.* $\overline{p}\gamma_\mu p\to\frac{1}{2}(\overline{p}\gamma_\mu p-\overline{p}\gamma_5\gamma_\mu\gamma_5 p)=\overline{p}\gamma_\mu p$, while $\overline{p}\gamma_\mu n\to\frac{1}{2}(\overline{p}(1-\gamma_5)n)$.

(*) A. SALAM: *Phys. Rev.* Letters (to be published). « γ_5-symmetrization » is not synonymous with « γ_5-invariance » of a Lagrangian though these may be equivalent in some special cases.

(**) If spin zero Bose fields are present $\varphi\to-\varphi$ simultaneously with $\psi\to\gamma_5\psi$. All currently accepted strong and weak Lagrangians appear to be γ_5-symmetrized.

It is now necessary to introduce strong interactions of baryons and mesons and to insure that these conserve charge in the conventional sense. This is an important point, for we do not wish to use the vector combinations $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ (and singlets B_{s1}, B_{s2}) etc., for writing strong interactions. If we did so, and if strong interactions were rotation invariant in the [3] Q space, there would be no distinction between these interactions and weak and e.m. interactions so far as conventional strangeness, isotopic-spin etc., are concerned. In other words it is necessary to state some criterion which guarantees that even if strong interactions are not rotation-invariant in charge-space these conserve charge in the conventional sense.

Such a criterion is easily stated. Among the $\boldsymbol{A}$ fields, γ_5 symmetry has picked out the component A^0. The divergence of the current $\mathcal{J}_\mu^{(1)}$ vanishes.

Here $\mathcal{J}_\mu^{(0)}$ equals the third component of

$$\varepsilon(\bar{\boldsymbol{L}}_1\boldsymbol{L}_1+\bar{\boldsymbol{L}}_1\boldsymbol{L}_2+\bar{\boldsymbol{B}}_1\boldsymbol{B}_1+\bar{\boldsymbol{B}}_2\boldsymbol{B}_2-\bar{\boldsymbol{B}}_1^c\boldsymbol{B}_1^c-\bar{\boldsymbol{B}}_2^c\boldsymbol{B}_2)+$$

$$+\text{terms from } \pi, \text{ K and } A \text{ fields. (See Equ. (10)).}$$

Thus charge operator

$$Q=\int \mathcal{J}_4^0 \mathrm{d}^3x\,,$$

can be defined in the usual manner. If we now demand that the hamiltonian for all strong interactions H_s must commute with Q (and must satisfy baryon conservation) charge in the conventional sense is conserved even if H_s is not rotation-invariant in Q-space. Thus the usual isotopic rotation invariant Lagrangian with 8 different coupling constants

$$g_1\bar{N}N\pi+g_2\bar{\Lambda}\Sigma\pi+g_3\bar{\Sigma}\Sigma\pi+\ldots\ g_5\bar{N}K\Lambda+g_6\bar{N}K\Sigma+\ldots$$

is fully acceptable. The mass degeneracy of baryons in the free Lagrangian will be removed by these strong interactions.

5. – The simple interaction Lagrangian written symbolically as

$$(\bar{\boldsymbol{L}}\boldsymbol{L}+\bar{\boldsymbol{B}}\boldsymbol{B}+\boldsymbol{\pi}\boldsymbol{\pi}+\boldsymbol{K}\boldsymbol{K}+\boldsymbol{A}\boldsymbol{A})\cdot\boldsymbol{A}$$

combines electro-magnetic interactions, as well as weak interactions. For weak interactions γ_5-symmetrization gives parity-violating terms while electromagnetic interaction terms still conserve parity. The weak interactions appropriately violate strangeness conservation while electromagnetic interactions do not.

There are also a number of conservation (eq. (10)) laws for the weak inter-

action currents. These are similar to the conservation laws previously discovered by GELL-MANN (*) and ZELDOVICH and GERŠTEJN.

The idea that weak interactions and electromagnetic interactions should be combined, originating as it did with SCHWINGER (2), has been often discussed privately, as has also the possibility that the resulting A-fields should be associated with a Yang-Mills type of gauge transformation. In this connection we are unable to disentangle the extent to which we are indebted to others. Particularly we would like to mention Dr. S. L. GLASHOW who has expressed similar ideas to us privately and a recent preprint (H. MAYER, Dubno) where a future paper on this subject is promised.

* * *

The work of one of the authors (J.C.W.) was supported by the U.S. Air Force through the Office of the Scientific Research of the Air Research and Development Command.

(*) M. GELL-MANN: *Phys. Rev.*, to be published. S. S. GERŠTEJN and ZELDOVICH: *Žurn. Èksper. Teor. Fiz.*, **29**, 698, (1958).

RIASSUNTO (*)

Il postulato di una « connessione locale » in un [3] spazio della carica conduce all'introduzione di tre campi di spin uno. Uno di questi può identificarsi col campo elettromagnetico e si può dimostrare che gli altri due sono il veicolo di tutte le interazioni deboli note, unificando così tali interazioni coll'elettromagnetismo. La teoria tien conto del fatto che, a differenza delle interazioni elettromagnetiche, le interazioni deboli violano la conservazione della parità e della stranezza.

(*) *Traduzione a cura della Redazione.*

$\Delta I = \frac{1}{2}$ Rule

ABDUS SALAM
Imperial College, London, England

and

J. C. WARD
Carnegie Institute for Technology, Pittsburgh, Pennsylvania
(Received September 20, 1960)

One attractive way of formulating the $\Delta I = \frac{1}{2}$ rule is to postulate the existence of "spurions"[1] — neutral isospinor particles carrying no energy and momentum which are emitted or absorbed in decays of strange particles. We wish to point out that conventional field theory contains a perfectly natural realization of spurions. Consider the decay mode $K_1^0 \to \pi^+ + \pi^-$. If the Lagrangian for the process is $f_\omega K_1^0 \pi^+ \pi^-$, it is easy to see that the vacuum expectation value of K_1^0 would not be zero. In fact $\langle K_1^0 \rangle \propto f_\omega$. If we now consider a seemingly isotopic-spin conserving T product of field operators, like $\Lambda N^\dagger \boldsymbol{\tau} \cdot \boldsymbol{\pi} K$, it not only possesses matrix elements for strong interactions like $N + \pi \to \Lambda + K$, but also a matrix element for the decay $\Lambda \to N + \pi$ (arising from $\Lambda N^\dagger \boldsymbol{\tau} \cdot \boldsymbol{\pi} \langle K \rangle$) consistent with the $\Delta I = \frac{1}{2}$ rule: $\langle K_1^0 \rangle$ thus gives a realization for the spurion.

One can perhaps reverse the thinking above and reinterpret isotopic-spin nonconservation in the following manner. Let us postulate that all fields possessing the same quantum numbers as the vacuum (and in particular K_1^0 for which $CP = +1$) do have nonzero expectation values. These expectation values define the so-called coupling constants (in this instance the weak coupling constant f_ω). Admitting only seemingly isotopic-spin conserving products of field operators, we obtain both isotopic-spin conserving and isotopic-spin nonconserving matrix elements from the same expression. Apparent nonconservation of isotopic spin is another expression for the fact that K_1^0 had nonzero expectation value.

[1] G. Wentzel, *Proceedings of the Sixth Annual Rochester Conference on High-Energy Nuclear Physics* (Interscience Publishers, New York, 1956).

Reprinted from Physical Review Letters, Vol. 5, No. 8, p. 390.
©1960 Amer. Inst. Phys.

It is possible to interpret parity nonconservation in the same manner. If there exist[2] K' mesons of opposite (strong) parity to K mesons, $\langle K_2'^0 \rangle \neq 0$. Thus $\lambda N^\dagger \boldsymbol{\tau} \cdot \boldsymbol{\pi}\gamma_5 K'$ gives a decay matrix element for $\Lambda \to N + \pi$ of opposite parity to one arising from $\Lambda N^\dagger \boldsymbol{\tau} \cdot \boldsymbol{\pi} K$, leading to parity nonconservation in the decay phenomenon.

Details of a theory of coupling constants and symmetry properties based on the above ideas will be published elsewhere.

We are grateful to Professor R. G. Sachs for the hospitality of the Summer Institute at the University of Wisconsin.

[2]M. Gell-Mann has re-emphasized the theoretical necessity of these particles for understanding weak interactions; Report at the Tenth Annual Rochester Conference on High-Energy Nuclear Physics, 1960 (to be published).

A. SALAM and J. C. WARD
1° Gennaio 1961
Il Nuovo Cimento
Serie X, Vol. 19, pag. 165-170
© 1961 Società Italiana di Fisica

On a Gauge Theory of Elementary Interactions.

A. SALAM
Imperial College - London

J. C. WARD
Carnegie Insitute of Technology - Pittsburgh

(ricevuto il 15 Settembre 1960)

Summary. — A theory of strong as well as weak interactions is proposed using the idea of having only such interactions which arise from generalized gauge transformations.

1. – Introduction.

One of the problems engaging current interest in field theory is the problem of determining which fields are « elementary » in some fundamental sense, and which are not. An equally, if not more, important, problem is that of finding a guiding principe for writing fundamental interactions of fields. The only such principle which exists at the present time seems to be the gauge-principle. Whenever a symmetry property exists, the associated gauge transformation leads in a definite manner to the postulation of an interaction through the mediation of a number of intermediate particles. There exist, at present, numerous attempts to understand all known elementary interactions in this manner. In an earlier paper [1], the authors considered a gauge-transformation in [3] « charge-space » to generate weak and electro-magnetic interactions. Recently J. J. SAKURAI [2] has used similar ideas to postulate five intermediate vector mesons which may be responsible for mediating strong interactions.

All these attempts suffer from certain weaknesses. Our earlier attempt [1] to understand weak and electro-magnetic interactions produced only parity-

[1] A. SALAM and J. C. WARD: *Nuovo Cimento*, 11, 568 (1959).
[2] J. SAKURAI: *Ann. of Phys.*, 11, 1 (1960).

conserving interactions in a natural manner. Besides, the weak interactions did not obey the $\Delta I = \frac{1}{2}$ rule. Sakurai's strong Lagrangian suffers from the defect that it contains no Yukawa-like terms permitting single emission of pions or K-mesons by baryons.

In this note we wish to reconsider the problem. Our basic postulate is that it should be possible to generate strong, weak and electro-magnetic interaction terms (with all their correct symmetry properties and also with clues regarding their relative strengths), by making local gauge transformations on the kinetic-energy terms in the free Lagrangian for all particles. This is the statement of an ideal which, in this paper at least, is only very partially realized.

It may however be of interest to set down the procedure which has been followed.

2. – A simple model.

Consider the π-nucleon system. Following SCHWINGER we assume the existence of a scalar iso-scalar particle σ. Write the free Lagrangian kinetic energy terms in the form

$$N_L^+ \gamma_4 \gamma_\mu \partial_\mu N_L + N_R^+ \gamma_4 \gamma_\mu \partial_\mu N_R + \tfrac{1}{2}(\partial_\mu \boldsymbol{\pi}) \cdot (\partial_\mu \boldsymbol{\pi}) + \tfrac{1}{2}(\partial_\mu \sigma)^2 , \tag{1}$$

where

$$N_L = \tfrac{1}{2}(1 + \gamma_5) N ,$$

$$N_R = \tfrac{1}{2}(1 - \gamma_5) N .$$

Following a suggestion made by SCHWINGER [3], GELL-MANN and LEVY [4] one may consider $\begin{pmatrix} N_L \\ N_R \end{pmatrix}$ as forming a spinor and $\begin{pmatrix} \sigma \\ \boldsymbol{\pi} \end{pmatrix}$ a 4-vector in a [4] Euclidean space. Since there are 6 rotations in such a space, the gauge-principle will give rise to six fields X, Y with the interactions

$$\begin{aligned} L_{\text{int}} = \frac{1}{2}\left(\sigma \frac{\partial \boldsymbol{\pi}}{\partial x_\mu} - \frac{\partial \sigma}{\partial x_\mu} \boldsymbol{\pi} + \boldsymbol{\pi} \wedge \frac{\partial \boldsymbol{\pi}}{\partial x_\mu} - \frac{\partial \boldsymbol{\pi}}{\partial x_\mu} \wedge \boldsymbol{\pi}\right) \cdot \mathbf{X}_\mu + \\ + i N_L^+ \gamma_4 \gamma_\mu \boldsymbol{\tau} \cdot \mathbf{X}_\mu N_L + \frac{1}{8}(\mathbf{X} \cdot \boldsymbol{\pi})^2 + \frac{1}{8}(\boldsymbol{\pi} \wedge \mathbf{X})^2 + \frac{1}{8} \mathbf{X}^2 \sigma^2 + \\ + \frac{1}{2}\left[\left(\sigma \frac{\partial \boldsymbol{\pi}}{\partial x_\mu} - \frac{\partial \sigma}{\partial x_\mu} \boldsymbol{\pi}\right) + \left(\frac{\partial \boldsymbol{\pi}}{\partial x_\mu} \wedge \boldsymbol{\pi} - \frac{\partial \boldsymbol{\pi}}{\partial x_\mu} \wedge \boldsymbol{\pi}\right)\right] \cdot \mathbf{Y}_\mu + \\ + i N_R^+ \gamma_4 \gamma_\mu \boldsymbol{\tau} \cdot \mathbf{Y}_\mu N_R + \frac{1}{8}(\mathbf{Y} \cdot \boldsymbol{\pi})^2 + \frac{1}{8}(\boldsymbol{\pi} \wedge \mathbf{Y})^2 + \frac{1}{8} \mathbf{Y}^2 \sigma^2 . \end{aligned} \tag{2}$$

---

[3] J. SCHWINGER: *Ann. of Phys.*, **2**, 407 (1957).

[4] M. GELL-MANN and M. LEVY: *Nuovo Cimento*, **14**, 705 (1960).

The « free » Lagrangians for the X field is as follows

$$\left[\left(\frac{\partial}{\partial x_\mu} X_\nu - X_\mu \wedge X_\nu\right) - \left(\frac{\partial}{\partial x_\nu} X_\mu - X_\nu \wedge X_\mu\right)\right]^2,$$

with a similar expression for the Y field.

One can rewrite the above interaction slightly differently, introducing fields

$$\tfrac{1}{2}(X + Y) = u\,,$$
$$\tfrac{1}{2}(X - Y) = v\,.$$

Thus

$$(3) \qquad L_{\text{int}} = \left[\left(\pi \wedge \frac{\partial \pi}{\partial x_\mu} - \frac{\partial \pi}{\partial x_\mu} \wedge \pi\right) + N^+ \gamma_4 \gamma_\mu \tau N\right] \cdot u + \\ + \left[\sigma \frac{\partial \pi}{\partial x_\mu} - \frac{\partial \sigma}{\partial x_\mu} \pi + N^+ \gamma_4 \gamma_\mu \gamma_5 \tau N\right] \cdot v + \text{terms quadratic in } u \text{ and } v\,.$$

We wish to identify the σ-containing part of the above Lagrangian as representing strong interactions and the remaining Lagrangian as representing weak interactions. The leptons (e^+, ν, e^-) or (μ^+, ν', μ^-) form a 3-vector in the space we are considering and a gauge-transformation will only link them with the u-field. The strength of the strong coupling comes about if we assume that the vacuum expectation value of σ ($\langle\sigma\rangle_0$) does not equal zero but equals $(g_s/2M_N)(1/g_\omega)$ (*).

Thus terms in the Lagrangian with $\sigma(\partial\pi/\partial x_\mu)\cdot v$ and $M^+\gamma_4\gamma_\mu\gamma_5\tau\cdot v N$ together give the conventional pseudo-vector strong Yukawa interaction with pions emitted singly. This is not to say that we are considering σ as an alternative expression for the strong coupling constant. There is every possibility that σ-particles are emitted (and absorbed) as physical particles.

Notice that the weak interactions in this model conserve parity. This unfortunate situation seems to persist in subsequent work also (**).

3. – Extension to strange particles.

In the above model σ, π form a 4-vector in a [4]-Euclidean manifold while N_L and N_R form a 4-spinor. A direct extension of this to include K-mesons is possible, provided we consider σ, π and K-particles to form a vector in an [8]-space while the sixteen baryons (eight-baryons each decomposed into their left and right components) form a 16-component spinor in such a space.

(*) Notice the terms $\sigma^2(X^2 + Y^2)$ in (2) could give the mass-terms for u and v particles.

(**) The theory of weak interactions recently proposed by GELL-MANN and LÉVY (4) effectively proceeds by identifying the terms containing X_μ only in (2) with the strangeness-conserving weak Lagrangian. The gauge-transformations giving rise to Y_μ fields are not considered.

The formalism we use was essentially developed by TIOMNO (5). Let us first recapitulate this. Write

$$\frac{1}{\sqrt{2}}(\Lambda^0 - \boldsymbol{\tau}\cdot\boldsymbol{\Sigma}) = \begin{pmatrix} Z^0 & \Sigma^+ \\ \Sigma^- & Y^0 \end{pmatrix} = (\Sigma_2, \Sigma_1)\,,$$

$$K^+ = K_1 - iK_2\,, \qquad K^0 = K_3 - iK_4\,.$$

If all K-coupling constants are equal, the conventional p.s. (or p.v.) strong K-Lagrangian can be written as

$$L_K = (N^+\Xi^+)(i\gamma_4\gamma_5)\begin{pmatrix} K^0 & K^+ \\ K^- & -\bar{K}^0 \end{pmatrix}\begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} + \text{h.c.}\,. \tag{4}$$

Write $\psi = \begin{pmatrix} N \\ \Xi \\ \Sigma_1 \\ \Sigma_2 \end{pmatrix}$; then L_K equals

$$\sum_{\alpha=1}^{4} \psi^+ i\gamma_4\gamma_5\Gamma_\alpha K_\alpha \psi\,, \tag{5}$$

where

$$\Gamma = \begin{pmatrix} & \tau\times 1 \\ \tau\times 1 & \end{pmatrix}, \qquad \Gamma_4 = \begin{pmatrix} & i\times 1 \\ -i\times 1 & \end{pmatrix}. \tag{6}$$

These Γ matrices are 8×8 matrices pertaining to a (six or) seven dimensional manifold. The spinor ψ is an 8×1 column.

It is possible to write the conventional p.s. (or p.v.) π-interactions in term of ψ.

Define three additional matrices,

$$\Gamma_{5,6,7} = \begin{pmatrix} 1\times\tau & & & \\ & 1\times\tau & & \\ & & -1\times\tau & \\ & & & -1\times\tau \end{pmatrix} = \tau_3\times 1\times\tau\,. \tag{7}$$

Then the matrices $\Gamma_1, \Gamma_2, \ldots, \Gamma_7$ anti-commute. If all π-couplings are equal (and in particular if $g_{\mu\mathcal{N}} = -g_{\pi\Xi\Sigma}$), one can write the π-Lagrangian

$$L_\pi = \sum_{\alpha=5,6,7} \psi^+ i\gamma_4\gamma_5(\Gamma_\alpha \pi_\alpha)\psi\,. \tag{8}$$

It is clear that π's and K's form a vector in a [7]-space.

So much for Tiomno's formalism. We can now follow a procedure analogous to Section 2 and obtain an expression which would contain terms like

$$\sigma\frac{\partial\pi}{\partial x_\mu}\cdot V_\mu\,, \qquad \sigma\frac{\partial K_\alpha}{\partial x_\mu}Z_\alpha\,, \qquad \sum_{\alpha=1}^{4}\psi^+(i\gamma_4\gamma_\mu\gamma_5)\Gamma_\alpha Z_\alpha\psi\,,$$

(5) J. TIOMNO: *Nuovo Cimento*, 6, 69 (1957).

to give an effective strong p.v. Lagrangian of the Tiomno type. Write

$$L_f = \psi_L^+ \gamma_4 \gamma_\mu \partial_\mu \psi_L + \psi_R^+ \gamma_4 \gamma_\mu \partial_\mu \psi_R + \tfrac{1}{2}[\partial_\mu \sigma + (\partial_\mu \pi)^2 + (\partial_\mu K_\alpha)^2]\,. \tag{9}$$

The sixteen-component entity $\begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$ forms a spinor in [8]-space. One anti-commuting Dirac set for such a space is

$$\begin{cases} \Gamma^{(8)}_{1,2,3} = \tau_1 \times (\tau_1 \times \boldsymbol{\tau}) \times 1\,, \\ \Gamma^{(8)}_{4} \;\; = \tau_1 \times (\tau_2 \times 1) \times 1\,, \\ \Gamma^{(8)}_{5,6,7} = \tau_1 \times (\tau_3 \times 1) \times \boldsymbol{\tau}\,, \\ \Gamma^{8}_{8} \;\; = \tau_2 \times 1 \times 1 \times 1\,. \end{cases} \tag{10}$$

In [8] space there are 28 rotations. Seven of these rotations ($\sigma \to \sigma + \boldsymbol{\epsilon}\cdot\boldsymbol{\pi} + \varepsilon_\alpha K_\alpha$, $\boldsymbol{\pi} \to \boldsymbol{\pi} - \boldsymbol{\epsilon}\sigma$, etc.) with the corresponding spinor rotation matrices

$$\frac{1}{2i}(\Gamma^{(8)}_8 \Gamma^8_\alpha - \Gamma^{(8)}_\alpha \Gamma^{(8)}_8)\,,$$

give

$$L_{\text{int}} = \frac{1}{2}\left(\sigma \frac{\partial \boldsymbol{\pi}}{\partial x_\mu}\cdot \boldsymbol{v}_\mu + \sigma \frac{\partial K_\alpha}{\partial x} Z_\alpha - \frac{\partial \sigma}{\partial x_\mu}\boldsymbol{\pi}\cdot\boldsymbol{v} - \frac{\partial \sigma}{\partial x_\mu} K_\alpha Z_\alpha\right) \tag{11}$$

i.e. Tiomno Lagrangian with $(i\gamma_4\gamma_5 \boldsymbol{\tau}\cdot\boldsymbol{\pi})$ replaced by $(i\gamma_4\gamma_\mu\gamma_5 \boldsymbol{\tau}\cdot\boldsymbol{v}_\mu)$ and $(i\gamma_4\gamma_5\Gamma_\alpha K_\alpha)$ replaced by $i\gamma_4\gamma_\mu\gamma_5\Gamma_\alpha Z_{\alpha,\mu}$. The fields $\boldsymbol{V}_\mu$ and Z_μ behave, so far as isotopic spin, etc., is concerned just like $\boldsymbol{\pi}$ and K mesons.

In so far as these seven rotation matrices do not form a Lie-Algebra, the interaction Lagrangian must contain 21 other fields corresponding to the remaining 21 rotations. From the point of view adopted in Section 2, these give weak interactions only. A general analysis of these terms has been given by GÜRSEY (*) in a recent paper which also adopts the Tiomno formalism to give an analogue of the Gell-Mann–Lévy theory of weak interactions.

It may be more profitable from our point of view to consider two fields σ and σ' in such a way that $(\sigma, \boldsymbol{\pi})$ form a 4-vector and (σ', K_α) a 5-vector. The resulting strong Lagrangian would then contain two coupling parameters $\langle\sigma\rangle_0$ and $\langle\sigma'\rangle_0$. $(\sigma, \boldsymbol{\pi})$ and (σ', K_α) spaces in a sense form two (disconnected) pieces of a [9] space. Even for the Tiomno Lagrangian it was possible to consider $\boldsymbol{\pi}$ and K-mesons as particles corresponding to disjunct pieces of the 7-dimensional space. Thus if we replace $\Gamma_\alpha \pi_\alpha$ in eq. (8) by $\Gamma'_\alpha \pi_\alpha$ where $\Gamma'_\alpha = 1 \times 1 \times \boldsymbol{\tau}$ (*i.e.* $g_{\pi\mathcal{N}} = + g_{\pi\Sigma\Sigma}$) we see that $\boldsymbol{\pi}$ and K no longer form a 7-vector. The per-

(*) F. GÜRSEY: preprint.

mitted rotations have to be limited in such a way that π-mesons are not transformed into K-mesons.

Returning to the [9] space, if (σ, π) form a [4] subspace and (σ', K) a [5] sub-space, it is clear that the total number of intermediate bosons will be $6+10=16$. Of these seven will mediate strong (p.v.) interaction, and 9 will mediate weak (v.) interactions. All these interactions conserve parity and isotopic-spin.

4. - $\Delta|I|=\frac{1}{2}$ rule.

One simple way to introduce strangeness violation consistent with the $\Delta|I|=\frac{1}{2}$ rule is to remark that the other field besides σ (or σ') which can have a non-zero expectation-value is the field corresponding to the θ^0 particle ($CP=+I$). In the notation above $\langle K_4\rangle \neq 0$ and would in fact be proportional to g_w even in the conventional theory. Thus all strangeness conserving terms like $\bar{\Lambda}N\tau\cdot\pi K$ describe also the matrix-elements for (parity-conserving) weak decay of $\Lambda\to\mathcal{N}+\pi$ consistent with $\Delta T=\frac{1}{2}$ provided we take the vacuum expectation value of the K-meson.

The non-zero expectation value of θ^0 is the perfect realization of the spurion idea of Wentzel so that it may not be necessary to introduce any additional fields to violate strangeness.

From what has been said above, it is clear that seven fields (three with transformation character of π-mesons and four with that of K-mesons) are necessary to mediate strong-interactions. The number of those necessary for weak interactions depends on the model used. However in all this work parity-violation for weak interactions remains a complete mystery.

* * *

We are deeply indebted to Professor R. G. SACHS for an invitation to the University of Wisconsin Summer Institute where part of the above work was completed.

RIASSUNTO (*)

Si propone una teoria delle interazioni sia forti che deboli nell'ipotesi di avere solo interazioni dovute a trasformazioni di gauge generalizzate.

(*) *Traduzione a cura della Redazione.*

A. SALAM and J. TIOMNO
16 Dicembre 1961
Il Nuovo Cimento
Serie X, Vol. 22, pag. 1287-1289
© 1961 Società Italiana di Fisica

Isotopic Spin Relations in Hyperon Production.

A. SALAM
Imperial College - London

J. TIOMNO
Centro Brasileiro de Pesquisas Físicas
Faculdade Nacional de Filosofia - Rio de Janeiro

(ricevuto il 12 Agosto 1961)

Experiments on Σ-K production on hydrogen for pion incident energies around 1 GeV seem to indicate that [1,2]

$$\sqrt{\sigma_+} + \sqrt{\sigma_-} \approx \sqrt{2\sigma_0}\,, \tag{1}$$

where:

$$\sigma_+ \to \pi_+ + p \to K_+ + \Sigma_+\,,$$
$$\sigma_0 \to \pi_- + p \to K_0 + \Sigma_0\,,$$
$$\sigma_- \to \pi_- + p \to K_+ + \Sigma_-\,.$$

Isotopic spin analysis would predict only an inequality whereas experiments seem to show an equality in (1).

The general amplitude $\pi + \mathcal{N} \to K + \Sigma$ can be expressed as

$$(\mathcal{N}\,|\,A\boldsymbol{\pi}\cdot\boldsymbol{\Sigma} + iB\boldsymbol{\tau}\cdot\boldsymbol{\pi}\wedge\boldsymbol{\Sigma}\,|\,K)\,. \tag{2}$$

In perturbation calculations A and B are real. Thus the experimental results can be understood in the form that the non-i-spin-flip amplitude $A \approx 0$ and only the i-spin-flip amplitude B contributes to the production process at these energies.

If $A \approx 0$ we have even a more constringent relation there (1):

$$\sigma_+ \approx \sigma_- \approx \tfrac{1}{2}\sigma_0\,. \tag{3}$$

Relation (3) is indeed favored by the experimental results [2] (Fig. 1). In the following we try to find the implications of this result on the basis of simple perturbation ideas.

Assuming even Σ-Λ and K-π parities we have three graphs given in Fig. 2.

---

(1) F. S. CRAWFORD, R. L. DOUGLASS, M. L. GOOD, G. R. KALBFLEISCH, M. L. STEVENSON and H. K. TICHO: *Phys. Rev. Lett.*, **3**, 394 (1959).

(2) C. BALTAY, H. COURANT, W. J. FICKINGER, E. C. FOWLER, H. L. KRAYBILL, J. SANDWEISS, J. R. SANFORD, D. L. STONEHILL and H. D. TAFT: *Rev. Mod. Phys.*, **33**, 374 (1961).

These give

$$(4)\quad \begin{cases} A \sim \left(\dfrac{g_{K\Sigma N} g_{N N\pi}}{p_N \cdot p_\pi} + \dfrac{g_{K\Lambda N} g_{\Lambda\Sigma\pi}}{p_\Sigma \cdot p_\pi} \right) (\gamma \cdot p_\pi) \\ B \sim \left(\dfrac{g_{K\Sigma N} g_{N N\pi}}{p_N \cdot p_\pi} - \dfrac{g_{K\Sigma N} g_{\Sigma\Sigma\pi}}{p_\Sigma \cdot p_\pi} \right) (\gamma \cdot p_\pi) \end{cases}$$

It is interesting to mention that relation (5) corresponds to Gell-Mann's D-coupling in the unitary theory [3], as

$$g_{K\Sigma N} = -\sqrt{3} g_{K\Lambda N};$$

$$g_{\Lambda\Sigma\pi} = 2/\sqrt{3} g_{N N\pi}.$$

The angular distribution is however roughly isotropic.

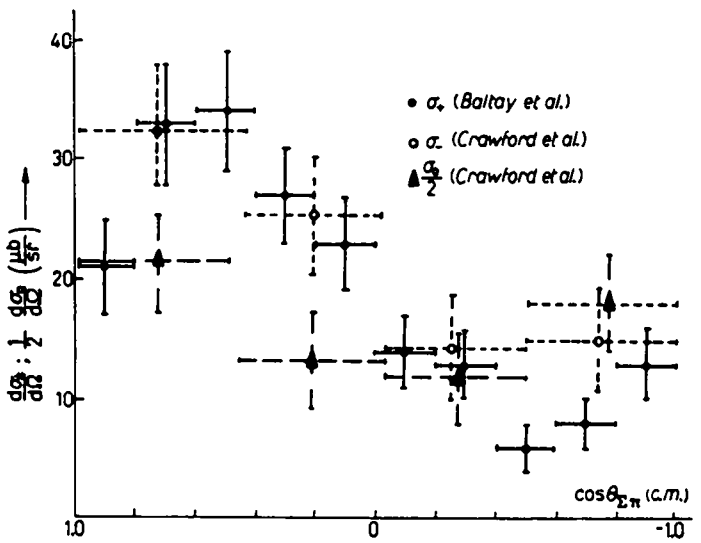


Fig. 1. – Angular distributions in $\pi_\pm + p \to K + \Sigma$ reaction at 1.08 GeV (Lab.). The Σ_0 cross-section is divided by 2.

where p's are the corresponding 4-momenta and we have neglected Λ-Σ mass difference. For the energy region considered $p_\Sigma \cdot p_\pi / p_N \cdot p_\pi \approx \frac{3}{2}$. Thus we may conclude that

$$(5)\qquad \frac{g_{K\Sigma N}}{g_{K\Lambda N}} \approx -\frac{3}{2}\frac{g_{\Lambda\Sigma\pi}}{g_{N N\pi}}.$$

In addition to the three graphs of Fig. 2 there is a fourth graph (Fig. 3), with a K* exchange [4] which contributes

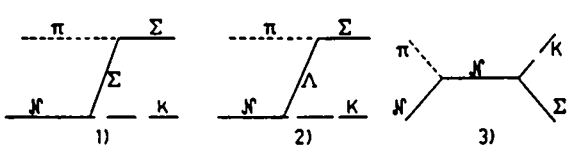


Fig. 2. – Feynman diagrams for Σ production.

[3] M. GELL-MANN: *The Eight-Fold Way*, California Institute of Technology report (1961).

[4] M. ALSTON, L. W. ALVAREZ, P. EBERHARD, M. L. GOOD, W. GRAZIANO, H. K. TICHO and S. G. WOJCICKI: *Phys. Rev. Lett.*, **5**, 520 (1960).

to both amplitudes A as well as B. If spin of K* is unity or if K* is scalar this will lead (as for K-Σ production ([5])) to a forward peaking of K-mesons. This

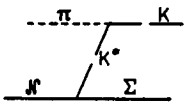


Fig. 3. – Diagram for Σ production with intermediate K*.

type of peaking does seem to show itself for a KΛ-production only at higher energies ($\gtrsim$ 1.3 GeV ([2-5])). This seems to indicate that $N\Sigma K^*$ coupling is energy-dependent and our neglect of this around 1 GeV may be justified.

When Λ-Σ parity is odd an K-π even the experimental relation (1) can be obtained only if the K* graph is no longer neglected. The relations of coupling constant are not any more simple. The angular distributions of K's may now be peaked backwards at $E_\pi \approx 1$ GeV, the forward peak due to the K* graph becoming important at higher energies, as indicated by the experimental results ([1,2,6]).

* * *

We wish to thank Prof. R. G. SACHS for the hospitality of the Summer Institute for Theoretical Physics.

([5]) M. GELL-MANN and J. TIOMNO: *Proc. of the 1960 Annual Intern. Conf. on High-Energy Physics at Rochester* (New York, 1960), p. 508; A. SALAM and J. C. WARD: *Phys. Rev. Lett.*, **5**, 390 (1960). Also J. TIOMNO, A. VIDEIRA and N. ZAGURY: *Phys. Rev. Lett.*, **6**, 120 (1961); M. BEG and P. DECELLES: *Phys. Rev. Lett.*, **6**, 145 (1961); C. CHAN: *Phys. Rev. Lett.*, **6**, 383 (1961).

([6]) M. I. SOLOVIEV: *Proc. of the 1960 Rochester Conference*, p. 388.

Reprinted from THE PHYSICAL REVIEW, Vol. 127, No. 3, 965–970, August 1, 1962

Broken Symmetries*

JEFFREY GOLDSTONE
Trinity College, Cambridge University, Cambridge, England

AND

ABDUS SALAM AND STEVEN WEINBERG†
Imperial College of Science and Technology, London, England

(Received March 16, 1962)

Some proofs are presented of Goldstone's conjecture, that if there is continuous symmetry transformation under which the Lagrangian is invariant, then either the vacuum state is also invariant under the transformation, or there must exist spinless particles of zero mass.

I. INTRODUCTION

IN the past few years several authors have developed an idea which might offer hope of understanding the broken symmetries that seem to be characteristic of elementary particle physics. Perhaps the fundamental Lagrangian is invariant under all symmetries, but the vacuum state[1] is not. It would then be impossible to prove the usual sort of symmetry relations among S-matrix elements, but enough symmetry might remain (perhaps at high energy) to be interesting.

But whenever this idea has been applied to specific models, there has appeared an intractable difficulty. For example, Nambu suggested that the Lagrangian might be invariant under a continuous chirality transformation $\psi \rightarrow \exp(i\theta \cdot \tau\gamma_5)\psi$ even if the fermion physical mass M were nonzero. But then there would be a conserved current J_λ, with matrix element

$$\langle p'|J_\lambda|p\rangle = f(q^2)\bar{u}'\gamma_5[i\gamma_\lambda - (2M/q^2)q_\lambda]u,$$

where $q=p-p'$. The pole at $q^2=0$ can only arise from a spinless particle of mass zero, which almost certainly does not exist. Of course, the pole would not occur if $f(0)=0$, which might be the case if we do not insist on identifying J_λ with the axial vector current of β decay. But Nambu showed that this unwanted massless "pion" also appears as a solution of the approximate Bethe-Salpeter equation.[1]

Goldstone[2] has examined another model, in which the manifestation of "broken" symmetry was the nonzero vacuum expectation value of a boson field. (This was suggested as an explanation of the $\Delta I=\frac{1}{2}$ rule by Salam and Ward.)[3] Here again there appeared a spinless particle of zero mass. Goldstone was led to conjecture that this will always happen whenever a continuous symmetry group leaves the Lagrangian but not the vacuum invariant.

* This research was supported in part by the U. S. Air Force under a contract monitored by the Air Force Office of Scientific Research of the Air Development Command and the Office of Naval Research.

† Alfred P. Sloan Foundation Fellow; Permanent address: University of California, Berkeley, California.

[1] Y. Nambu and G. Jona-Lasinio, Phys. Rev. 122, 345 (1961); W. Heisenberg, Z. Naturforsch. 14, 441 (1959).

[2] J. Goldstone, Nuovo cimento 19, 154 (1961).

[3] A. Salam and J. C. Ward, Phys. Rev. Letters 5, 512 (1960).

We will present here three proofs of this result. The first uses perturbation theory; the other two are much more general.

II. PERTURBATION THEORY

We will consider a multiplet of n spinless fields ϕ_i which interact among themselves and perhaps also with other fields. The Lagrangian is assumed to be invariant under a set of infinitesimal transformations:

$$\delta^\alpha\phi_i=\epsilon T_{ij}{}^\alpha\phi_j. \tag{1}$$

If the vacuum state were also invariant under these transformations, the vacuum expectation values of the ϕ_i would be subject to a set of linear relations,

$$T_{ij}{}^\alpha\langle\phi_j\rangle_0=0. \tag{2}$$

(Usually, there would be enough such relations to imply that all ϕ_i have zero vacuum expectation value. This is the case in the example to be discussed at the end of this section, where the ϕ_i transform as the defining representation of the orthogonal group, so that the T span the space of all antisymmetric matrices.)

We are going to examine the possibility that the vacuum state is not invariant under these transformations; in particular we will consider the consequences that ensue if

$$T_{ij}{}^\alpha\langle\phi_j\rangle_0\neq0 \tag{3}$$

for some α and some i. It is inconvenient to work with fields with nonzero vacuum expectation value, we we will define

$$\phi_i=\chi_i+\eta_i, \tag{4}$$

where

$$\eta_i\equiv\langle\phi_i\rangle_0,$$

so that χ_i is a quantum field with

$$\langle\chi_i\rangle_0\equiv0. \tag{5}$$

In perturbation theory this means that we should ignore all "tadpole" diagrams with a single external χ line.

The Lagrangian $L(\phi)$, although invariant under (1), will, in general, not be invariant under the "naive" transformations

$$\delta^\alpha\chi_i=\epsilon T_{ij}{}^\alpha\chi_j. \tag{6}$$

Hence, the vanishing of $\langle\chi_i\rangle_0$ provides a nontrivial self-consistency condition which allows us to calculate η_i up to some unavoidable ambiguities. We will now show that the value of η is such that propagators of some of the χ_i have a pole at zero mass.

We begin by defining a function $F(\eta)$ as the sum of all proper connected graphs with no external lines, and with the over-all energy momentum conservation factor $i(2\pi)^4\delta^4(0)$ omitted. Every factor λ_i in each term of $F(\lambda)$ represents a place where we might instead have an external line of type i. This can be seen in general by noting that the interaction Lagrangian density used in calculating these graphs is

$$L'(\chi,\eta)=L(\chi+\eta)-L_0(\chi),$$
$$L_0(\chi)=-\tfrac{1}{2}(\partial_\mu\chi_i)(\partial^\mu\chi_i)-\tfrac{1}{2}m^2\chi_i\chi_i. \tag{7}$$

It follows then that the sum $F^{(N)}$ of all connected proper diagrams with N external lines $i, j, \cdots$ carrying zero energy and momentum is (for $N\neq2$)

$$F_{ij\ldots}{}^{(N)}=(\partial^N/\partial\eta_i\partial\eta_j\cdots)F(\eta). \tag{8}$$

For $N=2$ the mass term in $-L_0$ gives an additional contribution, so

$$F_{ij}{}^{(2)}=(\partial^2/\partial\eta_i\partial\eta_j)F(\eta)+m^2\delta_{ij}. \tag{9}$$

As defined here, $F^{(N)}$ does not include the propagators for its external lines or the over-all factor $i(2\pi)^4\delta(0)$.

It is clear from the definition of $F^{(1)}$ that

$$F_i{}^{(1)}=(\Delta'^{-1}(0))_{ij}\langle\chi_j\rangle_0. \tag{10}$$

Here $\Delta'(p)$ is the complete propagator given by

$$\Delta'_{ij}(p)=\int d^4x\, e^{-ip\cdot x}\langle T\{\chi_i(x),\chi_j(0)\}\rangle_0. \tag{11}$$

Because $\langle\chi_i\rangle_0$ vanishes, the most general improper diagram for $\Delta'(p)$ can be constructed by stringing together proper self-energy parts $\Pi^*(p)$ into a linear chain; hence the inverse of the propagator is given as usual by

$$(\Delta'^{-1}(p))_{ij}=(p^2+m^2)\delta_{ij}-\Pi^*_{ij}(p). \tag{12}$$

For zero momentum,

$$\Pi^*_{ij}(0)=F_{ij}{}^{(2)}, \tag{13}$$

and so, using (9), we have

$$(\Delta'^{-1})_{ij}=-(\partial^2/\partial\eta_i\partial\eta_j)F(\eta). \tag{14}$$

We are going to prove that (14) has no inverse, so it should be kept in mind that it is $\Delta'^{-1}(0)$ and not $\Delta'(0)$ that is well defined.

To complete the proof, we now must make use of the invariance of $L(\phi)$ under the transformation (1). This has the consequence that it is only the presence of the η terms that breaks invariance under (6), and hence that $F(\eta)$ is invariant under the corresponding transformations

$$\delta^\alpha\eta_i=\epsilon T_{ij}{}^\alpha\eta_j. \tag{15}$$

Thus

$$(\partial F/\partial\eta_i)T_{ij}{}^\alpha\eta_j=0. \tag{16}$$

Differentiating with respect to η, this gives

$$(\partial^2F/\partial\eta_i\partial\eta_j)T_{ik}{}^\alpha\eta_k+(\partial F/\partial\eta_i)T_{ij}{}^\alpha=0. \tag{17}$$

For physically allowed values of η this relation together with (10) and (14), yields at zero momentum

$$(\Delta'^{-1})_{ij}T_{ik}{}^\alpha\eta_k=0. \tag{18}$$

We see that for zero momentum the inverse of the propagator becomes singular and so some elements of the propagator become infinite. This does not prove that there is a pole at zero mass, but we certainly expect the propagator to be infinite at $P^2=0$ only if the theory involves particles of zero mass. The fields with nonvanishing matrix element between the vacuum and states of zero mass are

$$\chi^a \equiv T_{ik}{}^a \eta_k \chi_i. \tag{19}$$

Clearly, none of this trouble would occur if it were not for our assumption (3) that $T_{ik}{}^a \eta_k \neq 0$. It is the broken symmetry, and not merely the nonzero vacuum expectation value η, that necessitates massless bosons.[4]

To see how this all works in a specific example, let us take as the Lagrangian

$$L(\phi) = -\bar{\psi}(\gamma\partial + M)\psi - \tfrac{1}{2}(\partial_\mu\phi_i)(\partial^\mu\phi_i) - \tfrac{1}{2}m^2\phi_i\phi_i - g\bar{\psi}O_i\psi\phi_i - \tfrac{1}{4}\lambda(\phi_i\phi_i)^2. \tag{20}$$

Defining $\phi_i = \chi_i + \eta_i$, this becomes

$$L(\phi) = L(\chi) - m^2(\chi_i\eta_i) - \tfrac{1}{2}m^2(\eta_i\eta_i) - g\bar{\psi}O_i\psi\eta_i - \lambda(\chi_i\chi_i)(\chi_j\eta_j) - \lambda(\chi_i\eta_i)^2 - \tfrac{1}{2}\lambda(\chi_i\chi_i)(\eta_j\eta_j) - \lambda(\eta_i\eta_i)(\chi_j\eta_j) - \tfrac{1}{4}\lambda(\eta_i\eta_i)^2. \tag{21}$$

The low-order contributions to the sum of all proper connected vacuum graphs are

$$\begin{aligned} F(\eta) = {} & F(0) - \tfrac{1}{2}m^2(\eta_i\eta_i) - \tfrac{1}{4}\lambda(\eta_i\eta_i)^2 + i(2\pi)^{-4}g\eta_i \\ & \times \int d^4p\ \mathrm{Tr}\{O_iS(p)\} + i(2\pi)^{-4}\lambda(\eta_i\eta_i + \tfrac{1}{2}\delta_{ij}\eta_k\eta_k) \\ & \times \int d^4p\ \Delta_{ij}(p) + \frac{i}{2}(2\pi)^{-4}g^2\eta_i\eta_j \int d^4p \\ & \times \mathrm{Tr}\{O_iS(p)O_jS(p)\} - i(2\pi)^{-4}\lambda^2(\eta_i\eta_j + \tfrac{1}{2}\delta_{ij}\eta_k\eta_k) \\ & \times (\eta_a\eta_b + \tfrac{1}{2}\delta_{ab}\eta_c\eta_c) \int d^4p\ \Delta_{ia}(p)\Delta_{jb}(p) \\ & - (2\pi)^{-8}\lambda^2\delta_{ij}\eta_k\delta_{ab}\eta_c \int d^4p\ d^4p' \\ & \times \{\Delta_{ia}(p)\Delta_{jb}(p')\Delta_{kc}(p+p') + \Delta_{ic}(p)\Delta_{jb}(p') \\ & \times \Delta_{ka}(p+p') + \Delta_{ib}(p)\Delta_{ja}(p')\Delta_{ka}(p+p')\}, \end{aligned} \tag{22}$$

where

$$S(p) = [-ip_\mu\gamma^\mu + M]^{-1},$$
$$\Delta_{ij}(p) = \delta_{ij}(p^2+m^2)^{-1}.$$

It can be readily seen that the derivative of $F(\eta)$ with respect to η_i is the sum $F_i^{(1)}$ of the proper connected diagrams with one external line, and that the second derivative with respect to η_i and η_j is the sum of the proper connected diagrams with two external lines, except that the $m^2\eta^2$ term in $F(\eta)$ does not contribute to $F^{(2)}$.

We will now assume that $L(\phi)$ is invariant under the group $SO(n)$ of orthogonal transformations on ϕ. This implies that the operators O_i are such that

$$\int d^4p\ \mathrm{Tr}\{O_iS(p)\} = 0, \tag{23}$$

$$\int d^4p\ \mathrm{Tr}\{O_iS(p)O_jS(p)\} = \delta_{ij}I. \tag{24}$$

Thus $F(\eta)$ is a function of $\eta^2 = (\eta_i\eta_i)$,

$$\begin{aligned} F(\eta^2) = {} & F(0) - \tfrac{1}{2}m^2\eta^2 - \tfrac{1}{4}\lambda\eta^4 \\ & + i(2\pi)^{-4}\eta^2(1+\tfrac{1}{2}n)\int d^4p\,(p^2+m^2)^{-1} \\ & + \tfrac{1}{2}i(2\pi)^{-4}g^2\eta^2 I - i(2\pi)^{-4}\lambda^2\eta^4(2+\tfrac{1}{4}n) \\ & \times \int d^4p\,(p^2+m^2)^{-1}(p^2+m^2)^{-1} - (2\pi)^{-8}\lambda^2\eta^2(2+n) \\ & \times \int d^4p \int d^4p'\,(p^2+m^2)^{-1}(p'^2+m^2)^{-1} \\ & \times [(p+p')^2+m^2]^{-1}. \end{aligned} \tag{25}$$

The dependence on η^2 alone implies then that

$$F_i^{(1)} = 2\eta_i F', \tag{26}$$

$$F_{ij}^{(2)} = (m^2 + 2F')\delta_{ij} + 4\eta_i\eta_j F''. \tag{27}$$

(A prime denotes differentiation with respect to η^2.) We see that the contribution of the term $-\frac{1}{2}m^2\eta^2$ in $F^{(2)}$ is canceled by the term $m^2\delta_{ij}$ in (27) arising from $-L_0$.

We have shown that $F_i^{(1)}$ is proportional to $\langle\chi_i\rangle_0$ and so must vanish. This implies that either η is zero or it must satisfy the consistency condition:

$$F'(\eta^2) = 0. \tag{28}$$

Thus if η is not zero, it is determined up to an orthogonal transformation; we certainly could not expect a more unambiguous determination.[5]

[4] It is clear from (19) that the maximum number of zero-mass fields is L, the number of Lie generators. There may in special cases be fewer than L zero-mass fields if not all fields χ^a given by (19) are linearly independent. This happens for example when $T_{ik}{}^a$ correspond to the "tensor" representations of simple Lie groups. For this case $T_{ik}{}^a$ are antisymmetric for all three indices. Therefore $\eta_a\chi^a = \eta_a T_{ik}{}^a \eta_k \chi_i = 0$, and only $(L-1)$ of the fields χ^a are linearly independent. These results are unaltered even if we allow in the theory more than one set of scalar fields ϕ_i with nonzero vacuum expectation values. To take a concrete case, the spurion theory proposed by Salam and Ward (reference 3) to explain the $\Delta I = \frac{1}{2}$ rule rests on assuming $\langle K_1^0\rangle \neq 0$. This would mean that the three companion fields to K_1^0, i.e., K^-, K^+, and K_2^0, must possess zero masses.

[5] Basic to the entire self-consistency procedure is, of course, the conjecture that $F'(\eta^2) = 0$ does possess a root for real η. By considering classical field theories, Goldstone states that a real root would exist provided the bare mass for the ϕ_i fields is pure imaginary. It is interesting to note that if the Lagrangian (20) contains no $-\frac{1}{4}\lambda(\phi_i\phi_i)^2$ term, an application of Lehmann's mass theorem [H. Lehmann, Nuovo cimento 11, 342 (1945)] shows that Goldstone's condition [(bare mass)$^2 < 0$] can never be satisfied. If, however, the $-\frac{1}{4}\lambda(\phi_i\phi_i)^2$ term is present in the Lagrangian, Lehmann's theorem gives no indication of the sign of (bare mass)^2.

We have also shown that

$$\Delta_{ij}'^{-1}(0)=m^2\delta_{ij}-F_{ij}^{(2)}$$
$$=-2F'\delta_{ij}-4\eta_i\eta_jF''. \tag{29}$$

If $\eta=0$, there is no reason to expect that $F'=0$, so that $(\Delta'^{-1})_{ij}$ is the nonsingular matrix $-2F'\delta_{ij}$. But if $\eta\neq 0$, then for physically allowed values of η we must have $F'=0$ so,

$$(\Delta'^{-1}(0))_{ij}=-4\eta_i\eta_jF''. \tag{30}$$

This is certainly a singular matrix. In fact

$$(\Delta'^{-1}(0))_{ij}u_j=0 \tag{31}$$

for any u orthogonal to η. All such u can be expressed as

$$u_i=T_{ij}\eta_j, \tag{32}$$

by choosing T_{ij} as an appropriate antisymmetric matrix. We see then that the space of u's is precisely the space indicated by the general considerations above.

III. GENERAL PROOFS

If the Lagrangian is invariant under an n-dimensional set of infinitesimal transformations which transform a general field ϕ_a according to

$$\delta\phi_a=\epsilon T_{ab}{}^{\alpha}\phi_b, \tag{33}$$

then there will exist a set of conserved currents

$$J^{\mu\alpha}=i\frac{\partial L}{\partial(\partial_\mu\phi_a)}T_{ab}{}^{\alpha}\phi_b, \tag{34}$$

$$\partial_\mu J^{\mu\alpha}=0. \tag{35}$$

The usual proof of the conservation equations (35) makes use only of the invariance of the Lagrangian, and hence should not be affected by the noninvariance of the vacuum. Also, from the canonical commutation relations we always expect that

$$[Q^{\alpha},\phi_a]=T_{ab}{}^{\alpha}\phi_b, \tag{36}$$

where

$$Q^{\alpha}=\int d^3x\, J^{0\alpha}(x). \tag{37}$$

We will begin by assuming again that there exists a set of spinless fields ϕ_i transforming according to Eq. (1), i.e.,

$$[Q^{\alpha},\phi_i]=T_{ij}{}^{\alpha}\phi_j. \tag{38}$$

The fields ϕ_i need not be "fundamental" here; all our remarks will apply equally well if the ϕ_i are synthetic objects like $\bar{\psi}O_i\psi$.

We shall show that if the vacuum is not annihilated by Q^{α}, so that

$$T_{ij}{}^{\alpha}\langle\phi_j\rangle\neq 0, \tag{39}$$

then the theory must involve massless particles.

The place we will look for zero-mass singularities is in the vacuum expectation value of the commutator of $J^{\mu\alpha}$ with ϕ_i. Using the usual Lehmann-Källén arguments, this can be written

$$\langle[J^{\mu\alpha}(x),\phi_i(y)]\rangle_0=\partial^\mu\int dm^2\,\Delta(x-y,m^2)\rho_i{}^{\alpha}(m^2), \tag{40}$$

where Δ is the usual causal Green's function for mass m and

$$(\Box^2-m^2)\Delta(x-y,m^2)=0, \tag{41}$$

$$(2\pi)^{-3}p^\mu\theta(p^0)\rho_i{}^{\alpha}(-p^2)=-\sum\delta(p-p^n)\langle 0|J^{\mu\alpha}{}_{(0)}|n\rangle$$
$$\times\langle n|\phi_i(0)|0\rangle. \tag{42}$$

The current conservation condition (35) together with (40) and (41) implies that

$$m^2\rho_i{}^{\alpha}(m^2)=0, \tag{43}$$

and hence

$$\rho_i{}^{\alpha}(m^2)=N_i{}^{\alpha}\delta(m^2), \tag{44}$$

$$\langle[J^{\mu\alpha}(x),\phi_i(y)]\rangle_0=N_i{}^{\alpha}\partial^\mu D(x-y), \tag{45}$$

where D is Δ for $m=0$. We would normally expect no singularity in $\rho(m^2)$, or in other words, $N_i{}^{\alpha}=0$. (It is well known, for example, that the pion-decay matrix element would vanish if the axial vector current were conserved.) But because of (39) we can show that $N_i{}^{\alpha}\neq 0$. For

$$N_i{}^{\alpha}=\int d^3x\, N_i{}^{\alpha}\partial^0 D(x)$$
$$=\langle[Q^{\alpha},\phi_i(0)]\rangle_0$$
$$=T_{ij}{}^{\alpha}\langle\phi_j(0)\rangle_0\neq 0. \tag{46}$$

Thus the sum in (42) must include states of zero mass.

It perhaps does not necessarily follow from the noninvariance of the vacuum that there exists (or can be constructed) a set of spinless fields ϕ_i with $T_{ij}{}^{\alpha}\langle\phi_j\rangle_0\neq 0$. We therefore wish to offer a simple nonrigorous argument that if there were no massless particles in the theory, then we would have to conclude that

$$0=|\alpha\rangle\equiv Q^{\alpha}|0\rangle; \tag{47}$$

for the conservation of current implies that Q^{α} and hence $|\alpha\rangle$ is invariant under the inhomogeneous Lorentz group. But then

$$\langle\alpha|J^{\alpha\mu}(x)|0\rangle=0, \tag{48}$$

so that

$$\langle\alpha|\alpha\rangle=\int d^3x\,\langle\alpha|J^{\alpha 0}(x)|0\rangle=0. \tag{49}$$

Equation (47) follows from (49) and the positive-definiteness assumption.

This "proof" is probably unobjectionable in ordinary theories with no massless particles. But if there are massless particles, the integrals in (37) and (49) become

somewhat poorly defined, because then there are states that are not Lorentz invariant but arbitrarily close to Lorentz invariance. If $|\alpha\rangle$ is such a state, the matrix element (48) will be small, but may give a large value to the integral (49).

As an example, let $|\mathbf{p}\rangle$ be a state containing a particle of mass zero, and construct the wave packet

$$|f\rangle=\int d^3p\ |\mathbf{p}|^{-\frac{1}{2}}f(|\mathbf{p}|)|\mathbf{p}\rangle. \tag{50}$$

The normalization condition is

$$1=4\pi\int|f(E)|^2EdE, \tag{51}$$

so that if we wish we can choose $f(0)\neq 0$. Lorentz invariance requires that

$$\langle\mathbf{p}|J^{\mu\alpha}(x)|0\rangle=N^\alpha|\mathbf{p}|^{-\frac{1}{2}}p^\mu e^{-ip\cdot x}, \tag{52}$$

where $p^0=|\mathbf{p}|$.

For particles with mass, current conservation would require that $N^\alpha=0$, but no such conclusion can be drawn for massless particles. For the wave packet (52), we now have

$$\langle f|J^{\mu\alpha}(x)|0\rangle=N^\alpha\int d^3p(p^\mu/|p|)f^*(|\mathbf{p}|)e^{-ip\cdot x},$$

so that

$$\langle f|\alpha\rangle=\int d^3x\,\langle f|J^{\mu\alpha}(x)|0\rangle$$

$$=(2\pi)^3N^\alpha f^*(0).$$

If we choose $f(0)$ to be nonzero, then the factor $|\mathbf{p}|^{-\frac{1}{2}}$ in (50) gives $|f\rangle$ a more-or-less Lorentz-invariant component, which has nonzero matrix element with $|\alpha\rangle$.

This becomes a bit more understandable if we ask ourselves what is the meaning of the state

$$(\exp i\epsilon Q^\alpha)\ |0\rangle=|0\rangle+i\epsilon\ |\alpha\rangle.$$

Clearly, this state is degenerate with $|0\rangle$, and is in fact another possible vacuum. It involves an infinitesimal component containing massless bosons of preponderantly low momentum. The role of the massless particles is apparently just to give meaning to the various possible vacua.

IV. PROSPECTS FOR THE UNSYMMETRIC VACUUM

The general proofs of the last section rest entirely on the assumption that there exists a conserved current, and that the integral of its time-component satisfies (38). This follows formally from the invariance of the Lagrangian, but in a *quantum* field theory the noncommutativity of the factors in the current, and the possible nonconvergence of the integral of its time-component, make our arguments essentially nonrigorous.

Therefore, it seems reasonable to defer belief in the necessity of massless bosons in a theory with unsymmetric vacua until such a *bête noire* is found in an actual calculation based on such a theory. We have already shown in Sec. II that the massless bosons do appear when we perform calculations using perturbation theory, provided that the symmetry of the theory is broken only by the choice of the vacuum expectation value η of the boson field.

But this is not the most general possibility. The original work of Nambu[1] indicates that the choice of a fermion mass can also break a symmetry. In this theory the fermion mass is

$$-ip^\mu\gamma_\mu=m_1+i\gamma_5m_2,$$

where (m_1,m_2) transform under chirality transformations like the components of a 2-vector. If m_1 and m_2 are not zero they must satisfy a condition of form

$$F(m_1^2+m_2^2)=0.$$

Any particular choice of direction for the vector (m_1,m_2) breaks the chirality invariance. (It should be noted that Nambu's choice $m_2=0$ is purely arbitrary and not dictated by parity conservation. For a general mass we must simply define the matrix associated with parity transformations to be

$$[(m_1+im_2\gamma_5)/(m_1^2+m_2^2)^{\frac{1}{2}}]\beta,$$

rather then just β.)

In Nambu's theory there is no "bare" spinless boson, but it is possible to construct a two-vector

$$\phi_1=\bar\psi\psi,\quad \phi_2=i\bar\psi\gamma_5\psi.$$

With Nambu's definition of parity (i.e., $m_2=0$) the vacuum expectation value of ϕ_2 but not of ϕ_1 vanishes, so the vector $\langle\boldsymbol{\phi}\rangle_0$ points in the 1-direction. An infinitesimal chirality transformation would rotate $\langle\boldsymbol{\phi}\rangle_0$ towards the 2-axis, so we are led to conjecture that the propagator of ϕ_2 has a zero-mass pole. In fact, just such a pole was found by Nambu in an approximate treatment of the bound-state problem. However, to show that the pole remains at zero mass when more complicated diagrams are considered would require a more thorough understanding of the treatment of bound states in perturbation theory. We are attempting this at present.

In a more complicated situation we could have an invariance broken both by the choice of a vacuum expectation value of a "bare" field and also simultaneously by the choice of a mass. For example, if we specialize the model discussed in Sec. II to the case of chirality invariance, we must take

$$M=0,\quad O_1=1,\quad O_2=i\gamma_5,$$

so that

$$L=-\bar\psi(\partial\cdot\gamma)\psi-\tfrac{1}{2}(\partial_\mu\phi_i)(\partial^\mu\phi_i)-\tfrac{1}{2}m^2(\phi_i\phi_i)$$
$$-g\bar\psi\psi\phi_1-ig\bar\psi\gamma_5\psi\phi_2-\tfrac{1}{4}\lambda(\phi_i\phi_i)^2.$$

In this case our conjecture would be that:

(1) If part of the loss of symmetry is due to the choice of a two-vector $\langle\phi\rangle_0$, then there must appear a zero-mass pole in the part of ϕ perpendicular to this vacuum expectation value.

(2) If part of the loss of symmetry is due to the choice of a nonzero Fermion mass $m_1+im_2\gamma_5$ then there must appear a zero-mass pole in the propagator of

$$\phi'=-m_2\bar{\psi}\psi+m_1\bar{\psi}\gamma_5\psi.$$

[Presumably this is the same pole as for (1). Parity conservation would require (m_1,m_2) to be in the direction of $\langle\phi\rangle_0$.]

(3) If part of the loss of symmetry is due to the choice of a noninvariant boson mass (i.e., if the residue of the pole at mass m in the propagator of ϕ_i and ϕ_j is a matrix which is not just a constant times δ_{ij}), then there must appear a *two-boson* pole at zero mass in the propagator of ϕ^2.

These "conjectures" can be taken as proved if we accept the arguments of Sec. III. We believe that we will also soon be able to prove these conjectures, in general, within the framework of perturbation theory.

If this is so, then there seem only three roads open to an understanding of broken symmetries based on the noninvariance of the vacuum:

(A) The particle interpretation of such theories might be revised (as in the Gupta-Bleuler method) so that the massless particles are not physically present in final states if they are absent in initial states. However, all our attempts in this direction have failed.

(B) The massless particles might really exist. The argument against this based on the Eötvös experiment might not apply if the particles carry quantum numbers, since then the scattering cross section of two macroscopic bodies due to exchange of the massless bosons would be proportional only to the numbers of atoms in each body and not (as for Coulomb forces or gravitation) to the squares of the numbers of atoms. But the couplings of these massless particles would presumably be quite strong, and would have shown up in exotic decay modes.

(C) Goldstone has already remarked that nothing seems to go wrong if it is just discrete symmetries that fail to leave the vacuum invariant. A more appealing possibility is that the "ur symmetry" broken by the vacuum involves an inextricable combination of gauge and space-time transformations.

Note added in proof. Recently, one of us (S. W., Proceedings of the 1962 Geneva Conference on High Energy Nuclear Physics) has developed a method of rewriting any Lagrangian in order to introduce fields for bound as well as "elementary" particles. This allows the proof of Sec. II to be extended to the case where the field with nonvanishing vacuum expectation value is any scalar function of the elementary particle fields, hence completing our argument.

ACKNOWLEDGMENTS

We would like to thank Dr. S. Kamefuchi for some helpful discussions. One of us (S. W.) would like to thank (A. S.) for his hospitality at Imperial College, and we all wish to gratefully acknowledge the hospitality of Professor R. G. Sachs at the 1961 Summer Institute of the University of Wisconsin.

Volume 13, number 2 PHYSICS LETTERS 15 November 1964

ELECTROMAGNETIC AND WEAK INTERACTIONS

A. SALAM and J. C. WARD *
Imperial College, London

Received 24 September 1964

One of the recurrent dreams in elementary particles physics is that of a possible fundamental synthesis between electro-magnetism and weak interactions [1]. The idea has its origin in the following shared characteristics:

1) Both forces affect equally all forms of matter-leptons as well as hadrons.
2) Both are vector in character.
3) Both (individually) possess universal coupling strengths. Since universality and vector character are features of a gauge-theory these shared characteristics suggest that weak forces just like the electromagnetic forces arise from a gauge principle.

There of course also are profound differences:

1) Electromagnetic coupling strength is vastly different from the weak. Quantitatively one may state it thus: if weak forces are assumed to have been mediated by intermediate bosons (W), the boson mass would have to equal 137 M_p, in order that the (dimensionless) weak coupling constant $g_w^2/4\pi$ equals $e^2/4\pi$.

In the sequel we assume just this. For the outrageous mass value itself ($M_w \approx 137\, M_p$) we can offer no explanation. We seek however for a synthesis in terms of a group structure such that the remaining differences, viz:

2) Contrasting space-time behaviour (V for electromagnetic versus V and A for weak).
3) And contrasting ΔS and ΔI behaviours both appear as aspects of the same fundamental symmetry. Naturally for hadrons at least the group structure must be compatible with SU_3.

Lepton interactions define both the unit of the electric charge and (from μ-decay) the (bare) value of weak coupling constant. Leptons therefore must be treated first.

There is only one genuine lepton multiplet (in the limit $m_e = m_\mu = 0$) which really treats the neutrino field on the same footing ** as μ and e. This is the Konopinski-Mahmoud multiplet.

$$L = \begin{pmatrix} \nu \\ e^- \\ \mu^+ \end{pmatrix} \tag{1}$$

In terms of SU_3 generators ***, the electric charge clearly equals:

$$Q_l = \begin{pmatrix} 0 & & \\ & -1 & \\ & & +1 \end{pmatrix} = -2U_3 = -\tfrac{2}{3}\sqrt{3}\,(I_0 - V_0) = 2(I_3 - V_3), \tag{2}$$

while the weak interaction (with no neutral currents) has the unique form †

* Permanent address, John Hopkins University, Baltimore.

** There are other schemes where one postulates multiplets consisting of a two-component neutrino field together with a four-component electron or muon. These do not satisfy even the most elementary requirement of a genuine group-structure, i.e. that in some limit at least, the particles concerned should be transformable, one into the other.

*** $[T^i, T^j] = i f^{ijk} T^k$

$\{T^i, T^j\} = \frac{1}{3}\delta^{ij}\,\mathbb{1} + d^{ijk}\,T^k$

$I_3 = T^3$, $I^\pm = \frac{1}{2}\sqrt{2}\,(T^1 \mp iT^2)$, $I_0 = T^8$, $[I_0, I] = 0$

$U_3 = \frac{1}{2}\sqrt{3}\,T^8 - \frac{1}{2}T^3$, $U^\pm = \frac{1}{2}\sqrt{2}\,(T^6 \mp iT^7)$,

$U_0 = \frac{1}{2}\sqrt{3}\,T^3 + \frac{1}{2}T^8$

$V_3 = \frac{1}{2}\sqrt{3}\,T^8 + \frac{1}{2}T^3$, $V^\pm = \frac{1}{2}\sqrt{2}\,(T^4 \mp iT^5)$,

$V_0 = \frac{1}{2}\sqrt{3}\,T^3 - \frac{1}{2}T^8$

Note

$Q_h = T^3 + \frac{1}{3}\sqrt{3}\,T^8 = \frac{2}{3}\sqrt{3}\;\; U_0 = \frac{2}{3}\sqrt{3}\,(I_0 + V_0) = \frac{2}{3}(I_3 + V_3)$

$Q_l = T^3 - \sqrt{3}\,T^8 = -2U_3 = -\frac{2}{3}\sqrt{3}\,(I_0 - V_0) = +2\,(I_3 - V_3)$

Explicitly,

$$Q_h = \begin{pmatrix} \frac{2}{3} & & \\ & -\frac{1}{3} & \\ & & -\frac{1}{3} \end{pmatrix}, \quad Q_l = \begin{pmatrix} 0 & & \\ & -1 & \\ & & +1 \end{pmatrix}$$

Q_h is the conventional hadron charge operator, Q_l gives lepton-charge.

† Define

$$\psi_L = \tfrac{1}{2}(1+\gamma_5)\,\psi \qquad \psi_R = \tfrac{1}{2}(1-\gamma_5)\,\psi$$

$$(A^+B)_L = \tfrac{1}{2}A^+\gamma_4\gamma_\mu(1+\gamma_5)B$$

Reprinted with permission from *Phys. Lett.* 13 (1964) pp. 168–171
© 1964 Elsevier Science Publishers B.V.

$$\mathcal{L}_{weak} = [((e^-)^\dagger \nu)_L + (\nu^\dagger \mu)_R]\, W^- + \text{h.c.}$$

$$= (I_{1L} + V_{1R})\, W_1 + (I_{2L} - V_{2R})\, W_2 . \quad (3)$$

Here $W^\pm = \frac{1}{2}\sqrt{2}\,(W_1 \pm iW_2)$ and $I_{1L} = (\psi^+ I_1 \psi)_L$ etc. Now $I_{1L} + V_{1R}$, $I_{2L} - V_{2R}$, $I_{3L} - V_{3R}$ generate an SU_2 sub-group. Since two of the group generators $I_{1L} + V_{1R}$, $V_{2L} - V_{2R}$ give the weak currents, ideally one would have liked the neutral component [2] $(I_{3L} - V_{3R})$ to represent electromagnetism. This unfortunately is not the case. $I_{3L} - V_{3R}$ equals $\frac{1}{4}(Q_l + 3\gamma_5 Q_h)$ so that in addition to the "correct' electromagnetic current $\frac{1}{2}Q_l$ there appears also unwanted parity violating term $\gamma_5 Q_h$. We are therefore forced to extend the group structure as follows:

The free Lagrangian $(\psi^+ \partial\psi)$ is invariant for the following $SU_2 \otimes U_1$ transformation

$$L' = UL$$

where

$$U = \exp ie[(I_L + V'_R)\varepsilon - \sqrt{3}\,l(I_{0L} - V_{0R})\epsilon_0]. \quad (4)$$

Here l is an arbitrary constant determining the relative strength of the Abelian gauge $(I_{0L} - V_{0R})$ compared to the gauge $(I_L + V'_R)$ $(V' = V_1, -V_2, -V_3)$. Following the well-known procedure, construct $\mathcal{L}_{int}$ by replacing $(\partial\psi)$ in $\mathcal{L}_f$ with the co-variant derivative

$$\mathcal{D}L = \partial L + ie[(I_L + V'_R)\cdot W - \sqrt{3}\,l(I_{0L} - V_{0R})W_0]L \quad (5)$$

In terms of the two orthogonal combinations of fields,

$$A^0 = W_0 \cos\theta + W_3 \sin\theta$$

$$X^0 = -W_0 \sin\theta + W_3 \cos\theta$$

$$(l = \tan\theta)$$

one can write the neutral component of (4) in the form

$$(\cos\theta\, Q_X X^0 + \sin\theta\, Q_l A^0) \quad (6)$$

where $Q_X = \frac{1}{4}\sec^2\theta\,(3\gamma_5 Q_h + (2\cos 2\theta - 1)Q_l)$. The full interaction Lagrangian equals

$$\mathcal{L}_{int} = -ieL^+\gamma_4\gamma_\mu[(I^-_L + V'^-)\, W^+ + \text{h.c.} +$$

$$+ \sin\theta\, Q_l A^0 + \cos\theta\, Q_X X^0]L . \quad (7)$$

This Lagrangian has the following characteristics

1. The electromagnetic interaction $ie \sin\theta$ $(\bar{\mu}^+\mu^+ - \bar{e}\, e)$ is necessarily accompanied by a neutral current term $-ie\cos\theta\, Q_X X^0$. For the special case $\theta = 30^0$ $(3l^2 = 1)$ this neutral current conserves parity and is pure axial-vector, with the same sign of coupling for e^- and μ^+. For $\theta = 45^0$ $(l^2 = 1)$ the orthogonal fields A^0 and X^0 equal $\frac{1}{2}\sqrt{2}\,(W_3 \pm W_0)$ and the full neutral Lagrangian reads

$$\tfrac{1}{2}\sqrt{2}\, ie\,(\mu^+\mu - e^+e)A_0 \quad (8)$$

$$- \tfrac{1}{2}\sqrt{2}\, ie\,[\nu^+\gamma_5\nu - \tfrac{1}{2}e^+(1+\gamma_5)\, l + \tfrac{1}{2}\mu^+(1-\gamma_5)\mu]X_0 .$$

Within Lepton physics the only assumption one need make is that X^0 is not mass-less (in order that the neutral X^0 current does not contribute to what is experimentally called the electric charge). When we come to consider hadrons, the absence of neutral leptonic currents interacting with heavy particles however requires that we assume X^0 particles are at least as massive as W^+ or W^-. The appearance of the X^0 current is about the minimum price one must pay to achieve the synthesis of weak and electromagnetic interactions we are seeking.

Since the interaction of W_3 with W^+, W^- (including the coupling strength) is completely determined by the gauge principle, it is crucial to check that the electric charge carried by W^+, W^- identically equals the charge on μ^+ end e^-. Group-theoretically this is equivalent to making sure that the W's belong to the *same* group representation $(SU_2 \times U_1)$ as L itself.

Concretely, in accordance with the transformation properties of the W's, one may define the field strengths

$$W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - e\, W_\mu \times W_\nu \quad (9)$$

$$W_{0\mu\nu} = \partial_\mu W_{0\nu} - \partial_\nu W_{0\mu} . \quad (10)$$

The gauge-Lagrangian for the W fields equals

$$\mathcal{L}(W) = \tfrac{1}{4}(W_{\mu\nu}\cdot W_{\mu\nu} + W_{0\,\mu\nu}W_{0\mu\nu}) . \quad (11)$$

In terms of A^0 and X^0, rewrite (9) and (10) in the form

$$W^\pm_{\mu\nu} = \partial_\mu W^\pm_\nu - \partial_\nu W^\pm_\mu \mp ieW^\pm_\mu\,(\cos\theta\, X^0 + \sin\theta\, A^0)_\nu$$

$$\pm ie\, W^\pm_\nu\,(\cos\theta\, X^0 + \sin\theta\, A^0)_\mu$$

$$W^0_{\mu\nu} = \partial_\mu X^0 - \partial_\nu X^0_\mu - ie\cos\theta\,(W^-_\mu W^+_\nu - W^+_\mu W^-_\nu)$$

$$A^0_{\mu\nu} = \partial_\mu A^0_\nu - \partial_\nu A^0_\mu - ie\sin\theta\,(W^-_\mu W^+_\nu - W^+_\mu W^-_\nu).$$

Clearly,

$$\mathcal{L}(W) = \tfrac{1}{4}(A^0_{\mu\nu}A^0_{\mu\nu} + X^0_{\mu\nu}X^0_{\mu\nu} + W^+_{\mu\nu}W^-_{\mu\nu} + W^-_{\mu\nu}W^+_{\mu\nu}) \quad (12)$$

Comparing the coupling constants in (7) and (9), it is obvious that the dynamical charge carried by W^- and W^+ equals respectively the charge on e^- and μ^+ both in sign and magnitude.

Note that the magnetic moment of $W^\pm$ equals 2 Bohr magnetons irrespective of the value of θ.

For leptons

$$Q_l = T^3 - \sqrt{3}\, T^8$$

while for hadrons the Gell-Mann - Nishijima formula gives

$$Q_h = T^3 + \tfrac{1}{3}\sqrt{3}\, T^8 .$$

To correspond to this fundamental difference, clearly the appropriate infinitesimal unitary transformation for hadrons equals

$$U = 1 + i(\varepsilon \cdot I + \tfrac{1}{3}\sqrt{3}\, l\, I_0)$$

instead of the leptonic transformation

$$U = 1 + i(\varepsilon \cdot I - \sqrt{3}\, l\, I_0).$$

Now weak interactions for hadrons experimentally appear to exhibit the pattern

$$F^V + D^A$$

It was pointed out [3,4] in an earlier paper that this precisely is the consequence of assuming that the 9-fold of baryons B transforms as a representation of the group [5] structure $(SU_3)_L \otimes (SU_3)_R$. Specialising for weak and electromagnetic interactions to the sub-group $(SU_2 \times U_1)$ in place of the full SU_3 group, consider the double gauge transformation

$$B' = [\exp i\,(I_L + V_R)]\; B\; [\exp -i\,(I_R + V_L)]$$

where I, stands for,

$$I = I_1 \epsilon_1 + I_2 \epsilon_2 + I_3 \epsilon_3 + \tfrac{1}{3}\sqrt{3}\, l\, I_0 \epsilon_0$$

and likewise for V.

With the standard gauge procedure, this gives rise to the currents

$$-\tfrac{1}{2} B^+ \gamma_4 \gamma_\mu\, [I + V, B]$$

$$-\tfrac{1}{2} B^+ \gamma_4 \gamma_\mu \gamma_5\, \{I - V, B\}$$

The neutral components of $\mathcal{L}_{int}$ are

$$\mathcal{L}_{int} = -ie\, \sin\theta\, A^0\, \mathrm{Tr}\; B^+ [Q_h, B] A^0_\mu - ie\, \cos\theta\, X^0 J_X \qquad (13)$$

where

$$J_X = \tfrac{1}{4} \sec^2\theta\; \mathrm{Tr}\, B^+ ((1 + 2\cos 2\theta)[Q_h, B] - \gamma_5 \{Q, B\}). \qquad (14)$$

Once again, note that for $\theta = 60^\circ$ ($l^2 = 3$) the new neutral current is purely axial vector *. It is gratifying that relative electric (as well as weak) charges on baryons are the same in magnitude as on leptons. This is not true in any obvious manner for the new neutral charge appearing in J_X.

Summarising the full Lagrangian equals

$$\mathcal{L}_{weak} = -ie\, W^- \Big(((e^-)^+ \nu)_L + (\nu_\mu^+)_R + \tfrac{1}{2}\mathrm{Tr}\, B^+ [I^+ + V^+, B] + \tfrac{1}{2}\mathrm{Tr}\, B^+ \gamma_5 \{I^+ - V^+, B\} + (15) + \text{meson terms} \Big)$$

$$\mathcal{L}_{em} = -ie\, \sin\theta\, A^0 (p^+ p + \ldots + \mu\, \mu + \ldots) \qquad (16)$$

$$\mathcal{L}_{neutral} = -ie\, \cos\theta\, X^0\, (L^+ Q_X L + J_X) \qquad (17)$$

A number of problems remain

1) Even though the neutral X^0 current conserves strangeness and even though it is only just the lepton pairs e^+e^-, $\mu^+\mu^-$, $\nu^-\nu$ which make their appearance.

The upper limit established in the CERN neutrino experiment on $\nu + p \rightarrow \nu + p$ (1% of σ_{weak}) implies that the effective coupling of X^0 particles must be much smaller than $g_{W^\pm}$. We speculate that there possibly may exist a geometrical relation among the effective coupling constants like

$$e^2 g_{X^0}^2 = (g_{W^\pm}^2)^2 .$$

We have no convincing reasons for this relation. This however appears to be the most plausible conjecture for g_X^2, once we do accept that $g_{W^\pm}^2 \neq e^2$.

2) The weak hadron Lagrangian $[I^+ + V^+]_F + \gamma_5 \{I^+ - V^+\}_D$ shows no sign of the Cabibbo suppression for the strangeness-changing currents $V^\pm$; it also predicts the combination $(V+A)$ in $\epsilon^- \rightarrow n$ + leptons compared to $(V-A)$ for β-decay.

There are two distinct views about Cabibbo's phenomenological Lagrangian $[I^\pm \cos\theta + V^\pm \sin\theta]_F + \gamma_5 (a[I^\pm \cos\theta + V^\pm \sin\theta]_F + b\{I^\pm \cos\theta + V^\pm \sin\theta\}_D)$. One is that the Cabibbo suprression is a consequence of the symmetry-breaking mechanism operating in strong-interaction physics. The second view postulates that the suppression is intrinsic and results from a rotation of the weak relative to the strong i-spin and hypercharge axes. Now the most general charge and CP-conserving rotations one may consider are the following

* To incorporate the mesons, the natural assumption [3] appears to be to consider the P,S. particles M, together with a set of yet undiscovered scalar mesons M_2, to form a 9-fold $M = M_1 + iM_2$. The appropriate gauge transformation is

$$M' = \exp(iI)\, M \exp(-iV)$$

giving the gauge Lagrangian (see ref. 5, eq. (33) -

$$\mathcal{L} = -\tfrac{1}{2}\mathrm{Tr}\, [(\partial M_1 + \tfrac{1}{2}[I+V, M_1] - \tfrac{1}{2}\{I - V, M_2\})^2 + (\partial M_2 + \tfrac{1}{2}[I+V, M_2)] + \tfrac{1}{2}\{I - V, M_2\})^2$$

$$B'_L = X(\theta_1)\, B_L X^{-1}(\theta_2)$$
$$B'_R = X(\theta_3)\, B_R X^{-1}(\theta_4)$$

where

$$X = \exp(2i\theta\, T^7).$$

The net-effect of these is to replace I and V in (18) by the appropriate combinations of $I(\theta) = X^{-1}(\theta)\, I(\theta)$ and $V(\theta)$. For $I^{\pm}$, $V^{\pm}$, this may indeed give the Cabibbo suppression but for neutral currents (since $Q(\theta) = Q \cos 2\theta + \frac{1}{2} \sin 2\theta\, T^6$) any value of θ other than 0° or 90° inevitably gives rise to neutral strangeness-changing currents ($K^0 \to \mu^+ + \mu^-$). For such currents there appears to be no experimental evidence. The "rotation" view is therefore incompatible with the present theory.

In a subsequent paper we consider the problem of starting with $[I+V]_F + \gamma_5 \{I-V\}_D$, to show that the strong symmetry-breaking mechanisms do lead to a relative suppression of $[V^{\pm}]_F$ and $\gamma_5 \{V^{\pm}\}_D$ currents, as well as the appearance of $\gamma_5 [V^{\pm}]_F$ terms.

The authors would like to thank Prof. J. Prentki for a discussion.

References

1. A.Salam and J.C.Ward, Il Nuovo Cimento 11 (1959) 568;
S.Glashow, Nuclear Phys. 10 (1959) 103;
J.Schwinger, Ann. of Phys. 2 (1957) 407.
2. The group structure $I_L + V'_R$ for weak interactions was studied by R.Gatto (Il Nuovo Cimento 28 (1963) 567; and Y.Neeman, Il Nuovo Cimento 27 (1963) 567.
3. A.Salam and J.C.Ward, A gauge theory of elementary interactions, Phys, Rev., to be published
4. Y.Nambu and P.Freund, Phys. Rev. Letters 12 (1964) 714;
M.Gell-Mann, Physics Letters 12 (1964) 63;
M.Gell-Mann and Y.Neeman, to be published.
R.Makunda, R.E.Marshak and S.Okubu, to be published.
5. This group was introduced in elementary particle physics by A.Salam and J.C.Ward, Il Nuovo Cimento 27 (1961) 922;
M.Gell-Mann, Phys. Rev. 125 (1962) 1067;
and in slightly different form by R.E.Marshak and S.Okubo, 19 (1961) 1226.

* * * * *

Reprinted from THE PHYSICAL REVIEW, Vol. 138, No. 2B, B420–B423, 26 April 1965

Relativistic Structure of $SU(6)$

R. DELBOURGO, ABDUS SALAM, AND J. STRATHDEE
International Center for Theoretical Physics, Trieste, Italy
(Received 10 December 1964)

It is shown that a relativistic basis for invariance under $SU(6)$ exists only if the group structure is extended to $U^+(6)\otimes U^-(6)$ for any interaction terms. The notion of an inhomogeneous extension $U_w^+(6)\otimes U_w^-(6)$ is introduced. This extension leaves the kinetic-energy terms invariant, though it still does not provide a fully satisfactory theory.

1. INTRODUCTION

WE wish to examine in this note the relativistic basis[1] of recent generalizations of Wigner's supermultiplet theory[2] to elementary-particle physics. We start with the assumption that so far as the relativistic and internal-symmetry structures are concerned it is sufficient to start with an elementary multiplet of Dirac spinors—elementary in the sense that it corresponds to the fundamental representation of the internal symmetry concerned. More specifically, for the internal symmetry group $U(3)$, this fundamental representation corresponds to a Dirac set of three (Sakata-like) quarks. Our assumption then amounts to saying that so far as group theory is concerned all particles can be considered as composed of Dirac quarks. There are three questions to be studied:

(A) What is the structure of the "algebras" formed from the Dirac matrices and the internal symmetry generators T^i?

(B) For which of these "algebras" are the kinetic energy and the mass terms in a free Dirac Hamiltonian invariant?

(C) What types of interaction Hamiltonians (if any) are invariant for each "algebra"?

2. THE STRUCTURE OF THE COMBINED ALGEBRAS

Given a set of Hermitian Dirac matrices γ^A and internal symmetry generators[3] T^i, we note that

$$[\gamma^A T^i,\gamma^B T^j]=\tfrac{1}{2}\{\gamma^A,\gamma^B\}[T^i,T^j]+\tfrac{1}{2}[\gamma^A,\gamma^B]\{T^i,T^j\}\,.$$

For the fundamental (n-fold) representation of any unitary group $U(n)$, the $n\times n$ matrices T^i span the entire Hermitian basis and therefore both $[T^i,T^j]$ and $\{T^i,T^j\}$ are expressible as linear sums of the T^i's themselves. The same is trivially the case also with the full set of the 16 Dirac matrices. Specializing to $U(3)$ (i.e., for the nine matrices T^i, $i=0,1,\ldots,8$), it is clear from the above that the 144 matrices $\gamma^A T^i$ ($A=1,\cdots,16$; $i=0,1,\cdots,8$) in general provide the set of generators for a $U(12)$ structure.

It is easy to see from the results for the anticommutators of the Dirac γ's given in the Appendix that the general $U(12)$ group contains two $U(6)$ subgroups each generated by the 36 matrices

$$\begin{aligned} U^+(6):&\quad \tfrac{1}{2}(1+i\gamma_5)T^i,\quad \tfrac{1}{2}(1+i\gamma_5)\sigma_{\mu\nu}T^i;\\ U^-(6):&\quad \tfrac{1}{2}(1-i\gamma_5)T^i,\quad \tfrac{1}{2}(1-i\gamma_5)\sigma_{\mu\nu}T^i. \end{aligned} \tag{1}$$

The crucial remark is that since $(1+i\gamma_5)\sigma_{\mu\nu}$ is a set of antisymmetric self-dual matrices, *there are only three independent ones among these;* and likewise for $(1-i\gamma_5)\sigma_{\mu\nu}$. Clearly, a $U(6)$-invariant parity-conserving theory must necessarily possess $U^+\leftrightarrow U^-$ symmetry. $U^+(6)$ and $U^-(6)$ clearly are straightforward generalizations of $SU^+(2)$ and $SU^-(2)$—the two subgroups into which the (Euclidean) group of rotations in 4 dimensions splits.

The relevant matrices are unimodular for a Lorentz metric, and one must therefore first resort to the "unitary trick" of Weyl, i.e., go to a Euclidean metric, generalize $U^+(2)$, $U^-(2)$ to $U^+(6)$ and $U^-(6)$, and then pass back to the Lorentz metric. Adopting an obvious nomenclature we shall call $U^\pm(6)$ the *homogeneous symmetry group* in contrast to the *inhomogeneous* groups. We shall consider this in the next section.

"Algebras" of the second kind based on the inhomogeneous rather than the homogeneous Lorentz group are generated if we combine general relativistic "spin operators" with the T^i's. The "spin operators" are products of the Dirac matrices with momentum; one example is the set of the Pauli-Lubanski operator (which in the rest frame of a particle gives its intrinsic spin)

$$w_\mu=\tfrac{1}{4}\epsilon_{\mu\nu\rho\kappa}\sigma_{\nu\rho}p_\kappa\,,\quad (p_\mu w_\mu=0)\,. \tag{2}$$

Since

$$\begin{aligned} &[w_\mu,w_\nu]=i\epsilon_{\mu\nu\rho\kappa}p_\rho w_\kappa\,,\\ &\{w_\mu,w_\nu\}=\tfrac{1}{2}(p_\mu p_\nu-p^2g_{\mu\nu})\,, \end{aligned} \tag{3}$$

a new $U(6)$ algebra[4] is generated by the 36 quantities T^i, $w_\mu T^i$. Since w_κ (like $\sigma_{\mu\nu}$) commutes with γ_5, it is also possible to set up the groups $U_w^\pm(6)$ with the generators

$$U_w^\pm(6):\quad \tfrac{1}{2}(1\pm i\gamma_5)T^i,\quad \tfrac{1}{2}(1\pm i\gamma_5)w_\mu T^i.$$

Now the w's are only one example of the general

[1] A. Salam, Phys. Letters 13, 354 (1964); F. Gürsey and L. A. Radicati, Phys. Rev. Letters 13, 173 (1964); R. P. Feynman, M. Gell-Mann, and G. Zweig, Phys. Rev. Letters 13, 678 (1964). This paper appeared while this work was in progress.

[2] E. Wigner, Phys. Rev. 51, 106 (1937).

[3] For details see A. Salam, in *Lectures in Theoretical Physics, Boulder, Colorado, 1959* (Interscience Publishers, Inc., New York, 1960).

[4] Strictly speaking, this is only an algebra in the rest frame.

class of "spin operators." Another operator has been described by Calogero.[5] This is the tensor

$$w_{\alpha\beta}=-i\epsilon_{\mu\nu\alpha\beta}\gamma_5\gamma_\mu p_\nu,$$

$$w_{\alpha\beta}'=\tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}w_{\gamma\delta}=-i\gamma_5(\gamma_\alpha p_\beta-\gamma_\beta p_\alpha).$$

For a free Dirac particle, Calogero shows that the "even" components of the w_μ and $w_{\mu\nu}$ (in the Foldy-Wouthuysen sense) represent, respectively,

$$w_b=E\sigma_b{}^l+m\sigma_b{}^t, \quad w_0=\mathbf{p}\cdot\boldsymbol{\sigma}^l$$

$$\tfrac{1}{2}\epsilon_{abc}w_{bc}=m\sigma_a{}^l+E\sigma_a{}^t, \quad w_{a0}=\mathbf{p}\times\boldsymbol{\sigma}^t.$$

Here $\boldsymbol{\sigma}^l$ and $\boldsymbol{\sigma}^t$ are the longitudinal (along $\mathbf{p}$) and transverse components of spin $\boldsymbol{\sigma}$. In the rest frame, therefore, w_μ and $w_{\mu\nu}$ possess the same physical significance. In a future paper we hope to come back to the complete algebra of these spin operators.

3. INVARIANCE OF THE LAGRANGIANS

A. The Free Lagrangian

The Pauli-Lubanski operator w_μ and the Calogero operator $w_{\alpha\beta}$ possess the remarkable property that the Dirac operator $D=\gamma\cdot p-m$ commutes with them. Thus,

$$\mathfrak{L}_0=\bar\psi(\gamma\cdot p-m)\psi$$

is invariant for $U_w(6)$ (with generators T^i, $w_\mu T^i$). Likewise, defining $\psi_{L,R}=\frac{1}{2}(1\pm i\gamma_5)\psi$, the terms $\bar\psi_L\gamma\cdot p\psi_L$ and $\bar\psi_R\gamma\cdot p\psi_R$ are invariant for $U_w{}^\pm(6)$ (generators T^i, $w_\mu{}^\pm T^i$), respectively. This is of course not true of the mass term $m(\bar\psi_R\psi_L+\bar\psi_L\psi_R)$.

Consider now the algebras $U^\pm(6)$ generated by the T^i's and the Dirac matrices. The transformations[6]

$$\begin{aligned}\psi_L'&=(1+\tfrac{1}{2}i\alpha_\mu{}^i T^i\sigma_\mu+i\alpha^i T^i)\psi_L,\\ \psi_R'&=(1+\tfrac{1}{2}i\beta_\mu{}^i T^i\sigma_\mu+i\beta^i T^i)\psi_R,\end{aligned} \tag{4}$$

even when taken in conjunction with the Lorentz transformation

$$p_\mu'=p_\mu+\epsilon_\mu{}^\nu p_\nu$$

do not leave the free Dirac Lagrangian invariant.[7] In fact,

$$\delta(\bar\psi_L\gamma\cdot p\psi_L)=\alpha_{\mu\nu}{}^i\bar\psi_L T^i(\gamma_\mu p_\nu-\gamma_\nu p_\mu)\psi_L,$$

where the summation is to be carried from $i=1$ to 8. The conclusion therefore is that so far as the free Lagrangian is concerned, the group $U_w(6)$ is the only one which leaves the Lagrangian invariant; $U_w{}^\pm(6)$ leaves the kinetic-energy term unchanged but not the mass term, while for the covariant groups $U^\pm(6)$ (defined with Dirac matrices σ_μ, etc., rather than the spin matrix w_μ) the free Lagrangian possesses no specially desirable transformation character.

B. The Interaction Lagrangian

It is at this stage that our difficulties start. It has so far appeared impossible to construct an interaction Lagrangian involving a product of a finite number of field operators which is invariant for $U_w(6)$ or $U_w{}^\pm(6)$. It would seem therefore that if $U_w(6)$ is a relatively exact symmetry of nature, only at S-matrix type of theory can be constructed for it.

On the other hand, for the group structures $U^\pm(6)$, even though the free Lagrangian is not invariant, one can write invariant interaction terms. For example,[8] in the Euclidian sense (with $\bar\psi\equiv\psi^\dagger$), the interaction part of the parity-conserving Lagrangian

$$\begin{aligned}\bar\psi(\gamma\cdot p-m)\psi+(\bar\psi\gamma_\mu T^i\psi)(\bar\psi\gamma_\mu T^i\psi)\\ +(\bar\psi\gamma_\mu\gamma_5 T^i\psi)(\bar\psi\gamma_\mu\gamma_5 T^i\psi)\end{aligned} \tag{5}$$

is invariant for $U^+(6)\otimes U^-(6)$ transformations. There is a total of 72 currents, of which only the $SU(3)$ ninefold $\bar\psi\gamma_\mu T^i\psi$ is conserved. The divergences of these currents are listed in the Appendix.

4. INVARIANCE OF $SU^+(3)\otimes SU^-(3)$ UNDER $SU(6)$

The considerations of Sec. 3 leave us with a dilemma. What is the $U(6)$ group of Gürsey, Radicati, and Sakita, if it is not $U_w(6)$? If we are willing to give up covariance of the group structure (though of course not of the basic Lagrangian), a noncovariant subgroup of the structure $U^+(6)\otimes U^-(6)$ is provided by the 36 generators

$$T^i,\ \sigma_{a\beta}T^i \quad (i,j=0,1,\cdots,8;\quad a,b=1,2,3).$$

This structure coincides with the little group $U_w(6)$ for the rest frame $\mathbf{p}=0$. Thus, the set of transformations

$$\begin{aligned}&\psi'=(1+i\epsilon^j T^j+\tfrac{1}{2}i\epsilon_{ab}{}^j\sigma_{ab}T^j)\psi,\\ &p_0'=p_0,\ p_a'=p_a+\epsilon_{ab}{}^0 p_b \quad \text{(space-rotation)}\end{aligned} \tag{6}$$

leaves the interaction $(\bar\psi\gamma_\mu T^i\psi)(\bar\psi\gamma_\mu T^i\psi)+(\bar\psi\gamma_\mu\gamma_5 T^i\psi)\times(\bar\psi\gamma_\mu\gamma_5 T^i\psi)$ invariant as well, so the mass term as well as $\psi^\dagger p_0\psi$ is invariant, and the kinetic-energy term is changed by

$$\delta\mathfrak{L}=2\epsilon_{ab}{}^i\bar\psi\gamma_a p_b T^i\psi, \quad (i=1,\cdots,8).$$

The differential conservation law does not hold for the 24 currents

$$\bar\psi\gamma_\mu\sigma_{ab}T^i\psi \quad (i=1,\cdots,8;\quad a,b=1,2,3).$$

Note quite generally that $\delta\mathfrak{L}=i\sum\epsilon\partial_\mu j_\mu$.

Now the Lagrangian (5) postulated above is precisely the type of Lagrangian previously written down in a different connection—in connection with what has been

[5] F. Calogero, Nuovo Cimento **10**, 280 (1961).

[6] Note that the factors $\frac{1}{2}(1\pm i\gamma_5)$ which appeared in the generators $(1\pm i\gamma_5)T^i$, $(1\pm i\gamma_5)\sigma_\mu T^i$ in (1) are now incorporated in $\psi_{L,R}$ in writing (4).

[7] If we extended the notion of p_μ to a 72-dimensional vector $p_\mu{}^{ij\pm}$ transforming according to the (35,1)⊕(1,35) representation, then the invariance of the free Lagrangian would be restored.

[8] $\delta(\bar\psi_L\gamma_\mu T^i\psi_L)=\frac{1}{2}i\alpha_{\nu\lambda}{}^j\bar\psi_L$
$\times[i(g_{\mu\nu}\gamma_\lambda-g_{\mu\lambda}\gamma_\nu)d^{ijk}T^k-i\epsilon_{\mu\nu\lambda\rho}f^{ijk}\gamma_\rho T^k]\psi_L.$

called the $[SU(3)]_L \otimes [SU(3)]_R$ theory.[9] It is perhaps instructive to write the equations of the gauge version of this theory in detail as well as the transformations involved. We shall use the Lorentz metric and not adopt the trick of passing to the Euclidean space.

Start with the Lagrangian

$$\begin{aligned}\mathfrak{L} = \tfrac{1}{2} \sum_{1,2} [&\tfrac{1}{2} Z_{\mu\nu}{}^i (\partial_\mu Z_\nu{}^i - \partial_\nu Z_\mu{}^i + q f^{ijk} Z_\mu{}^j Z_\nu{}^k) \\ &- \tfrac{1}{4} Z_{\mu\nu}{}^i Z_{\mu\nu}{}^i - \tfrac{1}{2}\mu^2 Z_\mu{}^i Z_\mu{}^i] \\ &+ \bar{\psi}_L \gamma \cdot (p + q Z_1{}^i T^i) \psi_L \\ &+ \bar{\psi}_R \gamma \cdot (p + q Z_2{}^i T^i) \psi_R - m \bar{\psi} \psi , \end{aligned} \quad (7)$$

where $Z_{1\mu}{}^i$ and $Z_{2\mu}{}^i$ are 18 gauge vector fields which can be expressed as sums and differences of vector and axial vector fields

$$Z_1 = V + A , \quad Z_2 = V - A .$$

Now if we specialize to the Weyl representation of the γ matrices, the $U^+(6) \otimes U^-(6)$ transformation written in Eq. (4), we have

$$\begin{aligned} \psi_L{}' &= (1 + i\alpha^i T^i + i\boldsymbol{\alpha}^i \cdot \boldsymbol{\sigma} T^i) \psi_L , \\ \psi_R{}' &= (1 + i\beta^i T^i + i\boldsymbol{\beta}^i \cdot \boldsymbol{\sigma} T^i) \psi_R , \end{aligned} \quad (8)$$

or

$$\psi' = [1 + i(\epsilon^i + i\gamma_5 \eta^i) T^i + i\boldsymbol{\sigma} \cdot (\boldsymbol{\varepsilon}^i + i\gamma_5 \boldsymbol{\eta}^i) T^i] \psi ,$$

with $\alpha = \epsilon + i\eta$, $\beta = \epsilon - i\eta$. Then it is easy to check that the Yukawa-like interaction terms in (7) are invariant, provided

$$\begin{aligned} Z_{10}{}^{i\prime} &= Z_{10}{}^i + f^{ijk}(\alpha^j Z_{10}{}^k + \boldsymbol{\alpha}^j \cdot \mathbf{Z}_1{}^k) , \\ \mathbf{Z}_1{}^{i\prime} &= \mathbf{Z}_1{}^i + f^{ijk}(\alpha^j \mathbf{Z}_1{}^k + \boldsymbol{\alpha}^j Z_{10}{}^k) - d^{ijk} \boldsymbol{\alpha}^j \times \mathbf{Z}_1{}^k , \end{aligned} \quad (9)$$

$$\begin{aligned} Z_{20}{}^{i\prime} &= Z_{20}{}^i + f^{ijk}(\beta^j Z_{20}{}^k - \boldsymbol{\beta}^j \cdot \mathbf{Z}_2{}^k) , \\ \mathbf{Z}_2{}^{i\prime} &= \mathbf{Z}_2{}^i + f^{ijk}(\beta^j \mathbf{Z}_2{}^k - \boldsymbol{\beta}^j Z_{20}{}^k) - d^{ijk} \boldsymbol{\beta}^j \times \mathbf{Z}^k . \end{aligned} \quad (10)$$

However, the free Lagrangian changes by

$$\begin{aligned} \delta \mathfrak{L}_0 = &2 \bar{\psi}_L \boldsymbol{\alpha}^i \times \boldsymbol{\gamma} \cdot \mathbf{p} T^i \psi_L + 2 \bar{\psi}_R \boldsymbol{\beta}^i \times \boldsymbol{\gamma} \cdot \mathbf{p} T^i \psi_R \\ &+ im(\boldsymbol{\alpha}^i - \boldsymbol{\beta}^i) \cdot (\bar{\psi}_R \boldsymbol{\sigma} T^i \psi_L - \bar{\psi}_L \boldsymbol{\sigma} T^i \psi_R) \\ &+ \text{meson terms.} \end{aligned} \quad (11)$$

Clearly, the mass term $\bar{\psi}\psi$ is invariant only if the γ_5-containing part of the transformation vanishes ($\eta = 0$ or $\alpha = \beta$). Also as stated before, the kinetic energy term is at least invariant for the part of the transformation (8) corresponding to pure rotations, viz.,

$$\psi \to (1 + i\boldsymbol{\varepsilon}^i T^i + i\boldsymbol{\varepsilon}^0 \cdot \boldsymbol{\sigma}) \psi , \quad \mathbf{p} \to \mathbf{p} + \boldsymbol{\varepsilon}^0 \times \mathbf{p} .$$

Note also that $Z_\mu Z_\mu$ is invariant in the Euclidean sense, i.e., $\delta(Z_0{}^2 + \mathbf{Z}^2) = 0$. We have omitted writing the meson equivalents of the fermion kinetic energy in Eq. (11) for the sake of brevity.

5. CONCLUSIONS

To summarize the situation with respect to combining the Lorentz with the internal symmetry groups: We have succeeded in writing down a complete field-theoretic formalism provided we extend the algebra of the homogeneous, and not the inhomogeneous, Lorentz group. The operators $(1 \pm i\gamma_5)\sigma_{\mu\nu}$ correspond to the two independent (angular momentum) operators which generate the homogeneous Lorentz group.

The group structures $U^+(6) \otimes U^-(6)$ are a direct generalization from $SU^\pm(2)$ to $U^\pm(6)$ of spinors of each kind. However, we note that these generalizations do not leave the kinetic-energy terms invariant. The physically significant group with 36 generators is then the generalization of the homogeneous Lorentz group, this generalization consisting of 36 generators of space rotations σ and the unitary transformations, $T^i, \sigma T^i$. For the rest frame of a single particle, this group coincides with $U_w(6)$, the "little group" of Gürsey, Radicati, and Sakita.

Note added in proof. Subsequent to the writing of this paper, the authors have been able to show that a desirable covariant $U(6)$ theory is provided by the group structure $U(12)$ discussed in Sec. 2. The details are published in Proc. Roy. Soc. (London) 9 February 1965.

ACKNOWLEDGMENTS

The authors are deeply indebted to Professor P. T. Matthews and Dr. J. Charap for stimulating discussions, who have developed the $U^+(6) \otimes U^-(6)$ formalism independently. Since completion of this work, the authors have also seen a preprint by M. A. B. Bég and A. Pais on the subject and a reference therein to a preprint by Lee, Cornwall, Freund, and Bardakci.[10]

APPENDIX

We list here the commutators and anticommutators that arise in the algebra of $U(12)$. For the $U(3)$ spin part we have

$$[T^i, T^j] = i f^{ijk} T^k , \quad \{T^i, T^j\} = d^{ijk} T^k .$$

The f and d are the same as those of Gell-Mann[9] and $\lambda = \frac{1}{2} T$ establishes the correspondence with his notation.

For the Dirac algebra we use the 16 matrices

$$\gamma^A = 1, \ \gamma_\mu, \ \sigma_{\mu\nu} = \tfrac{1}{2} i [\gamma_\mu, \gamma_\nu], \ \sigma_{\mu 5} = i\gamma_\mu \gamma_5, \ \gamma_5$$

for which $\bar{\psi} \gamma^A \psi$ is real. Then listing the results,

$$\begin{aligned} &[1, \sigma_{\mu\nu}] = 0 , \\ &\{1, \sigma_{\mu\nu}\} = 2\sigma_{\mu\nu} , \\ &[\gamma_\lambda, \sigma_{\mu\nu}] = 2i(g_{\lambda\mu}\gamma_\nu - g_{\lambda\nu}\gamma_\mu) , \\ &\{\gamma_\lambda, \sigma_{\mu\nu}\} = -2\epsilon_{\lambda\mu\nu\rho}\sigma_{\rho 5} , \\ &[\sigma_{\kappa\lambda}, \sigma_{\mu\nu}] = 2i(g_{\kappa\nu}\sigma_{\lambda\mu} + g_{\lambda\mu}\sigma_{\kappa\nu} - g_{\kappa\mu}\sigma_{\lambda\nu} - g_{\lambda\nu}\sigma_{\kappa\mu}) , \\ &\{\sigma_{\kappa\lambda}, \sigma_{\mu\nu}\} = 2(g_{\kappa\mu}g_{\lambda\nu} - g_{\lambda\kappa}g_{\kappa\nu}) - 2\epsilon_{\kappa\lambda\mu\nu}\gamma_5 , \\ &[\sigma_{\lambda 5}, \sigma_{\mu\nu}] = 2i(g_{\lambda\mu}\sigma_{\nu 5} - g_{\lambda\nu}\sigma_{\mu 5}) , \\ &\{\sigma_{\lambda 5}, \sigma_{\mu\nu}\} = -2\epsilon_{\lambda\mu\nu\rho}\gamma_\rho , \\ &[\gamma_5, \sigma_{\mu\nu}] = 0 , \\ &\{\gamma_5, \sigma_{\mu\nu}\} = \epsilon_{\mu\nu\rho\lambda}\sigma_{\rho\lambda} . \end{aligned}$$

[9] A. Salam and J. C. Ward, Nuovo Cimento 19, 167 (1961) and Phys. Rev. 136, B763 (1964); M. Gell-Mann, Phys. Rev. 125, 1067 (1962); Physics 1, 63 (1964); P. G. Freund and Y. Nambu, Phys. Rev. Letters 12, 714 (1964).

[10] See for example K. Bardakci, J. M. Cornwall, P. G. O. Freund and B. W. Lee, Phys. Rev. Letters 13, 698 (1964); 14, 48 (1965). See also M. A. Bég and A. Pais, Phys. Rev. Letters 14, 51 (1965).

We also include here the currents which generate $U^+(6)\otimes U^-(6)$. Thus, under (4) we have

$$\delta\mathfrak{L}=\partial_\lambda\left[\frac{\partial\mathfrak{L}}{\partial(\partial_\lambda\psi_L)}\delta\psi_L+\frac{\partial\mathfrak{L}}{\partial(\partial_\lambda\psi_R)}\delta\psi_R\right]$$

$$+\text{meson contributions}$$

$$=-\partial_\lambda[\bar\psi_L\gamma_\lambda(\tfrac{1}{2}i\alpha_{\mu\nu}{}^j\sigma_{\mu\nu}T^j+i\alpha^jT^j)\psi_L$$

$$+\bar\psi_R\gamma_\lambda(\tfrac{1}{2}i\beta_{\mu\nu}{}^j\sigma_{\mu\nu}T^j+i\beta^jT^j)\psi_R+\cdots]$$

$$=-\tfrac{1}{2}i\alpha_{\mu\nu}{}^j\partial_\lambda J_{L\mu\nu,\lambda}{}^j-\tfrac{1}{2}i\beta_{\mu\nu}{}^j\partial_\lambda J_{R\mu\nu,\lambda}{}^j$$

$$-i\alpha^j\partial_\lambda J_{L\lambda}{}^j-i\beta^j\partial_\lambda J_{R\lambda}{}^j,$$

where

$$J_{L\mu\nu,\lambda}{}^j=\bar\psi_L\gamma_\lambda\sigma_{\mu\nu}T^j\psi_L+\cdots,\quad J_{L\lambda}{}^j=\bar\psi_L\gamma_\lambda T^j\psi_L+\cdots$$

and similarly for J_R. Since the only change in the Lagrangian is that of the free part,

$$\left.\begin{aligned}&\partial_\lambda J_{L\lambda}{}^j=\partial_\lambda J_{R\lambda}{}^j=0,\quad j=0,1,\cdots,8\\&\partial_\lambda J_{L\mu\nu,\lambda}{}^j=2\bar\psi_L(\gamma_\nu\partial_\mu-\gamma_\mu\partial_\nu)T^j\psi_L+\cdots\\&\partial_\lambda J_{R\mu\nu,\lambda}{}^j=2\bar\psi_R(\gamma_\nu\partial_\mu-\gamma_\mu\partial_\nu)T^j\psi_R+\cdots\end{aligned}\right\}j=1,\cdots,8$$

and

$$\partial_\lambda J_{R\mu\nu,\lambda}{}^0=\partial_\lambda J_{L\mu\nu,\lambda}{}^0=0\quad\text{if}\quad m=0.$$

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 284, *pp.* 146–158, 1965

The covariant theory of strong interaction symmetries

BY A. SALAM,† F.R.S.
Imperial College, London

AND R. DELBOURGO AND J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

(*Received* 18 *January* 1965)

A classification of particles is suggested based on a $\tilde{U}(12)$ symmetry scheme. This is a relativistic generalization of the $U(6)$ symmetry. The spin $\frac{1}{2}$ and $\frac{3}{2}$ baryons are each described by 20-component spinors which satisfy Bargmann–Wigner equations and belong to the 364 representation of the $\tilde{U}(12)$ group while the vector and p.s. mesons belong to the representation 143. The procedure for writing fully relativistic form factors is worked out in detail for baryon–meson and meson–meson cases.

The new results are the following:

(1) $\dfrac{F^C(q^2)}{F^M(q^2)} \propto 1+\dfrac{q^2}{\langle 2\mu\rangle m^2}$, where F^C and F^M are (Sachs) electromagnetic form factors.

(2) $\mu_p = 1+2m/\langle\mu\rangle$, where $\langle\mu\rangle$ is the mean mass of the 1^- multiplet and m the nucleon mass.

(3) $\mu_{\rho,K^*} = 3$.

The conventional $U(6)$ results can be recovered by projecting to the positive energy subspace in the rest system for each particle. To any irreducible representation of the $U(6)$ there corresponds one irreducible representation of $\tilde{U}(12)$ and vice versa.

1. INTRODUCTION

The problem of finding a relativistic generalization of the $U(6)$ group structure has engaged considerable attention recently (see references A). In an earlier paper (Delbourgo, Salam & Strathdee 1965*b*, to be referred to as I) it was suggested that one way to write relativistic S-matrix elements is to embed $U(6)$ in a $\tilde{U}(12)$ group-structure. The present paper gives the detailed formalism for writing relativistic S-matrix elements in this theory. In particular we compute the two basic baryon–meson and meson–meson form factors. The generalization of the symmetry from $U(6)$ to $\tilde{U}(12)$ gives the following new results:

(1) There is essentially just one *relativistic* form factor in strong interaction physics of the octet baryons and the (1^-) and (0^-) mesons. The relevance of this result to the conventional nucleon electric and magnetic form factors is discussed in §6.

(2) In Bohr magnetons the magnetic moments of the proton and the neutron are

$$\mu_p = 1+2x, \quad \mu_n = -\tfrac{2}{3}(1+2x),$$

where $x = m_N/\langle\mu\rangle$ and $\langle\mu\rangle$ is the mean mass of the (1^-) multiplet. The experimental magnetic moment values are well reproduced if $\langle\mu\rangle \approx 1000\,\text{MeV}$. This mass value is not far from the mean of m_ρ, m_ω, m_ϕ, etc.

† On leave of absence at the International Centre for Theoretical Physics, Trieste, Italy.

147 *The covariant theory of strong interaction symmetries*

The formalism describes both the spin $\frac{1}{2}$ and $\frac{3}{2}$ baryons as 20-component† composite entities made from the basic 4-component (Dirac) quark (Gell-Mann 1964; Zweig 1964). In §2 we describe the $\tilde{U}(12)$ algebra; in §3 are computed the form factors incorporating full $\tilde{U}(12)$ symmetry. In §4 we specialize to the reduced symmetry group $U(3) \times \mathscr{L}_4$ (where $\mathscr{L}_4$ refers to the homogeneous Lorentz group), the inhomogeneous Lorentz group ($I\mathscr{L}_4$) being considered in §5. The reduction of the $\tilde{U}(12)$ symmetry to $U(3) \times (I\mathscr{L}_4)$ in §§4 and 5 correctly reproduces the final physical symmetry situation, and gives the *relativistic* expressions for the baryon form factors in a fundamentally broken $\tilde{U}(12)$ symmetry scheme. The exact $U(6)$ symmetry can be recovered from our expressions in the limit of zero momenta.

2. $\tilde{U}(12)$ AND ITS SUBGROUPS

We assume that the fundamental entity for strong interactions is a 12-component (Dirac) quark. The group structure $\tilde{U}(12)$ is defined by the algebra of the 144 matrices $F^{Ri} = \gamma^R T^i$, $R = 1, \ldots, 16$; $i = 0, \ldots 8$. Here

$$\gamma^R = 1, \quad \gamma_\mu, \quad \sigma_{\mu\nu} = \tfrac{1}{2}\mathrm{i}[\gamma_\mu, \gamma_\nu], \quad \mathrm{i}\gamma_\mu\gamma_5, \quad \gamma_5,$$

with γ_0 hermitian and $\boldsymbol{\gamma}$ antihermitian and the metric $(1, -1, -1, -1)$. The general $\tilde{U}(12)$ transformation on the quark field $\psi_A = \psi_{p\alpha}$ ($p = 1, 2, 3$; $\alpha = 1, 2, 3, 4$) will be assumed to be

$$\delta\psi_{p\alpha} = \mathrm{i}(\epsilon^j + \epsilon^j_5\gamma_5 + \epsilon^j_\mu\gamma_\mu + \mathrm{i}\epsilon^j_{\mu 5}\gamma_\mu\gamma_5 + \tfrac{1}{2}\epsilon^j_{\mu\nu}\sigma_{\mu\nu})^\beta_\alpha (T^j)^q_p \psi_{q\beta}, \tag{2·1}$$

where all 144 ϵ's are real and this property leaves $\overline{\psi}\psi = \psi^+\gamma_0\psi$ invariant.

For higher representations of $\tilde{U}(12)$ (made up compositively from quarks) the transformation (2·1) will take the form

$$\delta\Psi = \mathrm{i}(\epsilon^j F^j + \epsilon^j_5 F^j_5 + \epsilon^j_\mu F^j_\mu + \epsilon^j_{\mu 5} F^j_{\mu 5} + \tfrac{1}{2}\epsilon^j_{\mu\nu} F^j_{\mu\nu})\Psi. \tag{2·2}$$

The general commutation rules of the generators F are listed in the appendix. The quadratic Casimir operator is

$$F^j F^j - F^j_5 F^j_5 + \tfrac{1}{2}F^j_{\mu\nu} F^j_{\mu\nu} + F^j_\mu F^j_\mu - F^j_{\mu 5} F^j_{\mu 5}.$$

Inspection of the commutators for $\tilde{U}(12)$ reveals that a 72-component subalgebra is generated by the operators $F^j, F^j_5, F^j_{\mu\nu}$. This is the subgroup $W(6)$ (see references A) and in the fundamental representation has the generators T^j, $\gamma_5 T^j$, and $\sigma_{\mu\nu} T^j$. The expressions

$$F^j_\mu F^j_\mu - F^j_{\mu 5} F^j_{\mu 5} \quad \text{and} \quad F^j F^j - F^j_5 F^j_5 + \tfrac{1}{2}F^j_{\mu\nu} F^j_{\mu\nu}$$

are now separately invariant under $W(6)$. Note that $W(6)$ possesses the important 36-parameter subgroup $U(6)$ ($T^i\sigma_{ab}$; T^i; $a, b = 1, 2, 3$). This will later be identified with the $U(6)$ of Gürsey, Radicati & Sakita.

† Of all the approaches to the relativistic $U(6)$ theory listed in references A, the formulation nearest in spirit to our own is that of P. Roman & J. J. Agasshi. The mesons are described similarly in both approaches but the baryon states (and their underlying symmetries) are different. Thus the well-known eightfold of baryons appears in the Agasshi–Roman theory with four components contrasted with the 20 components for our case. These essential differences come about because we use Bargmann–Wigner equations of §5 to make sure we are dealing with particles having a definite mass as well as definite spin.

3. Some representations of $\bar{U}(12)$ and their decomposition

A. The fundamental representation of $\bar{U}(12)$ is the 12-component quark discussed above. Following the usual procedure we assign to this quark the baryon number $B = \frac{1}{3}$. The baryons are then to be constructed from three quark states and the mesons from quark–antiquark states.

These states decompose under $\bar{U}(12)$ in the following way:

$$\left.\begin{aligned} \underline{12}\otimes\underline{12}^* &= \underline{1}+\underline{143}, \\ \underline{12}\otimes\underline{12}\otimes\underline{12} &= \underline{220}+\underline{364}+\underline{572}+\underline{572}. \end{aligned}\right\} \tag{3·1}$$

The $\underline{220}$ is completely antisymmetrical, the $\underline{364}$ completely symmetrical and the $\underline{572}$ is of the mixed symmetry type $[2, 1]$.

Under the subgroup $W(6)$ these states reduce according to

$$\left.\begin{aligned} \underline{143} &= (35,1)+(6,6^*)+(6^*,6)+(1,35)+(1,1), \\ \underline{220} &= (20,1)+(15,6)+(6,15)+(1,20), \\ \underline{364} &= (56,1)+(21,6)+(6,21)+(1,56), \\ \underline{572} &= (70,1)+(21,6)+(6,21)+(1,70). \end{aligned}\right\} \tag{3·2}$$

Under the subgroup $U(3)\otimes\bar{U}(4)$, where $U(3)$ refers to the space of unitary spin matrices T^j, and $\bar{U}(4)$ to that of the Dirac matrices γ^R, the contents are given by

$$\left.\begin{aligned} \underline{143} &= (8,15)+(1,15)+(8,1), \\ \underline{220} &= (8,20')+(10,\bar{4})+(1,20), \\ \underline{364} &= (10,20)+(8,20')+(1,\bar{4}), \\ \underline{572} &= (10,20')+(8,20)+(8,20')+(1,20')+(8,\bar{4}), \end{aligned}\right\} \tag{3·3}$$

where in each bracket the first number denotes the $U(3)$ representation and the second the $\bar{U}(4)$ representation. The three $\bar{U}(4)$ representations denoted here by 20, 20′ are of symmetry types [3] and [2, 1] respectively.

We note also the reduction of the products

$$\left.\begin{aligned} \underline{143}\otimes\underline{143} &= 1+\underline{143}_F+\underline{143}_D+\underline{4212}+\underline{5005}+\underline{5005}^*+\underline{5940}, \\ \underline{364}\otimes\underline{364}^* &= \underline{1}+\underline{143}+\underline{5940}+\underline{126412}, \\ \underline{143}\otimes\underline{364} &= \underline{364}+\underline{572}+\underline{16016}+\underline{35100}. \end{aligned}\right\} \tag{3·4}$$

B. Since it is our intention to assign the baryon to the $\underline{364}$, we compute the expectation values of the 144-vector $(\gamma^R T^j)^B_A$ between 364 states, namely

$$J^{Rj} = \bar{\Psi}^{ABC}(\gamma^R T^j)^{A'}_{A}\Psi_{A'BC}, \tag{3·5}$$

where Ψ_{ABC} is fully symmetric and has the $U(3)\times\bar{U}(4)$ decomposition

$$\Psi_{\alpha p\beta q\gamma r} = \frac{\sqrt{3}}{2\sqrt{2}}D_{\alpha\beta\gamma,pqr}+\epsilon_{pqr}V_{[\alpha\beta\gamma]}+\frac{1}{2\sqrt{6}}(\epsilon_{pqs}N_{[\alpha\beta]\gamma,r}^{\ \ s}+\epsilon_{qrs}N_{[\beta\gamma]\alpha,p}^{\ \ s}+\epsilon_{rps}N_{[\gamma\alpha]\beta,q}^{\ \ s}). \tag{3·6}$$

Here (α, β, γ) take the values 1, 2, 3, 4 and (p, q, r) the values, 1, 2, 3. $V_{[\alpha\beta\gamma]}$ is completely antisymmetric, $D_{\alpha\beta\gamma, pqr}$ is completely symmetric in both $\alpha\beta\gamma$ and pqr and $N_{[\alpha\beta]\gamma, r}^{\ \ s}$ is of the symmetry type [2, 1], i.e.

$$\left.\begin{aligned} N_{[\alpha\beta]\gamma} + N_{[\beta\alpha]\gamma} &= 0, \\ N_{[\alpha\beta]\gamma} + N_{[\beta\gamma]\alpha} + N_{[\gamma\alpha]\beta} &= 0. \end{aligned}\right\} \tag{3·7}$$

After some algebra we find (specializing to $i \neq 0$ so that $(T^i)_p^p = 0$),

$$\begin{aligned} J^{Ri} = {} & \tfrac{3}{8}\bar{D}^{\alpha\beta\gamma, pqr}(\gamma^R)_\alpha^{\alpha'}(T^i)_p^{p'} D_{\alpha'\beta\gamma, p'qr} \\ & + \tfrac{1}{4}[\bar{D}^{\alpha\beta\gamma}(\gamma^R)_\alpha^{\alpha'}(T^i)_p^{p'} \epsilon_{p'qs} N_{[\alpha'\beta]\gamma, r}^{\ \ s} + \bar{N}^{[\alpha\beta]\gamma, r}{}_s \epsilon^{pqs}(\gamma^R)_\alpha^{\alpha'}(T^i)_p^{p'} D_{\alpha'\beta\gamma, p'qr}] \\ & + [\bar{V}^{[\alpha\beta\gamma]}(\gamma^R)_\alpha^{\alpha'}(T^i)_p^{p'} N_{[\beta\gamma]\alpha', p'}^{\ \ \ \ p} + \bar{N}^{[\beta\gamma]\alpha, p}{}_{p'}(T^i)_p^{p'}(\gamma^R)_\alpha^{\alpha'} V_{[\alpha'\beta\gamma]}] \\ & - \tfrac{1}{24}(\bar{N}^{[\beta\alpha]\gamma}(\gamma^R)_\alpha^{\alpha'} N_{[\alpha'\beta]\gamma})_{3D+5F}^i \\ & + \tfrac{1}{12}(\bar{N}^{[\beta\alpha]\gamma}(\gamma^R)_\alpha^{\alpha'} N_{[\alpha'\gamma]\beta})_{3D+2F}^i, \end{aligned} \tag{3·8}$$

where

$$\left.\begin{aligned} (\bar{N}N)_F^i &= \bar{N}_r^p (T^i)_p^q N_q^r - \bar{N}_r^p N_p^q (T^i)_q^r, \\ (\bar{N}N)_D^i &= \bar{N}_r^p (T^i)_p^q N_q^r + \bar{N}_r^p N_p^q (T^i)_q^r. \end{aligned}\right\} \tag{3·9}$$

Notice the appearance at this stage of the characteristic combinations $3D + 2F$ and $3D + 5F$ for the form factors involving the eightfold baryons.

4. Reduction of $\bar{U}(4)$ to the homogeneous Lorentz group $\mathscr{L}_4$

A. We now specifically consider the space-time symmetries. So far it has been assumed that in its space-time behaviour the fundamental 4-component ($\bar{U}4$) entity ψ_α transforms as

$$\psi_\alpha \to S_\alpha^\beta \psi_\beta,$$

where

$$S_\alpha^\beta = 1 + \mathrm{i}(\epsilon_5 \gamma_5 + \epsilon_\mu \gamma_\mu + \mathrm{i}\epsilon_{\mu 5}\gamma_\mu\gamma_5 + \tfrac{1}{2}\epsilon_{\mu\nu}\sigma_{\mu\nu})_\alpha^\beta \tag{4·1}$$

with $\bar{\psi}$ transforming as $\bar{\psi}^\alpha \to \bar{\psi}^\beta [S^{-1}]_\beta^\alpha$. With ϵ's real, these transformations preserve the invariance of $\bar{\psi}^\alpha \psi_\alpha$. The higher representations $\Psi_{\alpha\beta\ldots}^{\gamma\delta\ldots}$ of $\bar{U}(4)$ transform as

$$\Psi_{\alpha\beta\ldots}^{\gamma\delta\ldots} \to S_\alpha^{\alpha'} S_\beta^{\beta'} \ldots (S^{-1})_{\gamma'}^\gamma (S^{-1})_{\delta'}^\delta \ldots \Psi_{\alpha'\beta'\ldots}^{\gamma'\delta'\ldots} \tag{4·2}$$

The symmetry represented by (4·1), however, is too general. In space-time terms it corresponds, as is well known, to the full conformal group symmetry C_4. To make contact, however, with physical space-time symmetries of (at this stage) the *homogeneous Lorentz group* we must descend from $\bar{U}(4)$ to $\mathscr{L}_4$. There are a number of ways of doing this which are not all necessarily equivalent so far as the underlying physics is concerned as will be discussed in §5.

Disregarding the problems connected with unitary spin, clearly the most direct symmetry reduction is achieved by taking

$$\epsilon_5 = \epsilon_\mu = \epsilon_{\mu 5} = 0.$$

Now as is well known for this case (*though not for* $\bar{U}(4)$) one can define an antisymmetric matrix $(C^{-1})^{\alpha\beta}$ (within the Dirac algebra) with the defining property that $C^{-1}\psi^T$ transforms similarly to $\bar{\psi}$. (Here ψ^T is the transpose of ψ.) In particular, $C^{-1}\psi^T\psi$ just like $\bar{\psi}\psi$ is an invariant. Clearly from the definition above, the

antisymmetric matrix $C^{-1} = (C^{-1})^{\alpha\beta} = -(C^{-1})^{\beta\alpha}$ plays the role for $\mathscr{L}_4$ of the metric tensor. We may regard $(C^{-1})^{\alpha\beta}$ as a contravariant quantity (with two upper indices) and its inverse $C_{\alpha\beta}$ as the corresponding covariant $(C_{\alpha\beta}(C^{-1})^{\beta\gamma} = \delta_\alpha^\gamma)$.

It is easy to show (Jauch & Rohrlich 1955) that the matrix C with the defining property above† can be realized by finding a matrix satisfying

$$(\gamma_\mu C)_{\alpha\beta} = (\gamma_\mu C)_{\beta\alpha},$$

where

$$(\gamma_\mu C)_{\alpha\beta} = (\gamma_\mu)_\alpha^\gamma C_{\gamma\beta}.$$

It is also easy to show that the 16 Dirac matrices $(\gamma^R C)_{\alpha\beta}$ fall into two distinct classes; the matrices $(\gamma_\mu C)_{\alpha\beta}$ and $(\sigma_{\mu\nu} C)_{\alpha\beta}$ are symmetric, and $C_{\alpha\beta}$, $(\gamma_5 C)_{\alpha\beta}$, $(\mathrm{i}\gamma_\mu\gamma_5 C)_{\alpha\beta}$ are antisymmetric. For writing symmetric and antisymmetric higher-rank 'spinors' in $\mathscr{L}_4$ these are the primary quantities one needs.

To illustrate consider the following examples:

(i) *Multi-spinor of rank* 2

A second-rank symmetric spinor must have the form

$$\Phi_{(\alpha\beta)} = [(\gamma_\mu C)\,\phi_\mu + \tfrac{1}{2}(\sigma_{\mu\nu} C)\,\phi_{\mu\nu}]_{\alpha\beta}. \tag{4·3}$$

Likewise the general antisymmetric spinor has the form

$$\Phi_{[\alpha\beta]} = [C\phi + (\gamma_5 C)\,\phi_5 + \mathrm{i}(\gamma_\mu\gamma_5 C)\,\phi_{\mu 5}]_{\alpha\beta}. \tag{4·4}$$

(ii) *Fully symmetric spinor of rank* 3

Consider $\Psi_{\alpha\beta\gamma}$ with full symmetry in α, β, γ. From symmetry in γ and β, one may write Ψ in the form

$$\Psi_{\alpha\beta\gamma} = \psi_{\alpha\mu}(\gamma_\mu C)_{\beta\gamma} + \tfrac{1}{2}\psi_{\alpha\mu\nu}(\sigma_{\mu\nu} C)_{\beta\gamma}. \tag{4·5}$$

We now show that full symmetry in α, β, γ is realized provided

$$\gamma_\mu \psi_\mu = 0, \tag{4·6}$$

$$\gamma_\mu \psi_{\mu\nu} + \mathrm{i}\psi_\nu = 0. \tag{4·7}$$

For, the three antisymmetric tensors $(C^{-1})^{\gamma\alpha}$, $(C^{-1}\gamma_5)^{\gamma\alpha}$ and $(\mathrm{i}C^{-1}\gamma_\mu\gamma_5)^{\gamma\alpha}$ must annihilate $\Psi_{\alpha\beta\gamma}$.

This gives

$$\psi_{\alpha\mu}(\gamma_\mu)_\beta^\alpha + \tfrac{1}{2}\psi_{\alpha\mu\nu}(\sigma_{\mu\nu})_\beta^\alpha = 0,$$

$$\psi_{\alpha\mu}(\gamma_\mu\gamma_5)_\beta^\alpha + \tfrac{1}{2}\psi_{\alpha\mu\nu}(\sigma_{\mu\nu}\gamma_5)_\beta^\alpha = 0,$$

$$\psi_{\alpha\mu}(\gamma_\mu\gamma_\lambda\gamma_5)_\beta^\alpha + \tfrac{1}{2}\psi_{\alpha\mu\nu}(\sigma_{\mu\nu}\gamma_\lambda\gamma_5)_\beta^\alpha = 0.$$

If we suppress Dirac indices, the first two equations give (4·6) and $\sigma_{\mu\nu}\psi_{\mu\nu} = 0$, and the last is equivalent to (4·7). Note that as a result of (4·6) and (4·7) the 40-component entity on the right side of (4·5) contains only 20 independent components.

† One could have made the transition $\tilde{U}(4) \to \mathscr{L}_4$ via the intermediate symmetry stage $\widetilde{Sp}(4)$ ($\epsilon_s = \epsilon_{\mu 5} = 0$). The $\widetilde{Sp}(4)$ transformation with 10 parameters (ϵ_μ, $\epsilon_{\mu\nu}$) also admits of the existence of this matrix C. All statements of §4 apply equally to theories with $\widetilde{Sp}(4)$ symmetry.

(iii) *Mixed spinor of rank* 3

The 20-component tensor $\Psi_{[\alpha\beta]\gamma}$ which satisfies the 'trace' condition†

$$\Psi_{[\alpha\beta]\gamma}+\Psi_{[\gamma\alpha]\beta}+\Psi_{[\beta\gamma]\alpha}=0 \tag{4·8}$$

can necessarily be written in the form

$$\Psi_{[\alpha\beta]\gamma}=(\gamma_5 C)_{\alpha\beta}\psi_\gamma+\mathrm{i}(\gamma_\mu\gamma_5 C)_{\alpha\beta}\psi_\gamma+C_{\alpha\beta}K_\gamma. \tag{4·9}$$

The mixed symmetry character of Ψ yields the constraint

$$\gamma_5\psi+\mathrm{i}\gamma_\mu\gamma_5\psi_\mu-K=0. \tag{4·10}$$

This follows on multiplying (4·8) by $(C^{-1})^{\alpha\beta}$.

B. To return to $\tilde{U}(12)$, we now decompose all irreducible $\tilde{U}(12)$ higher representations relative to the representations of $U(3)\otimes\mathscr{L}_4$, maintaining the over-all symmetry. Thus for the $(8\times 20'')$ part of the $\underline{364}$ representation, we write

$$N_{[\alpha\beta]\gamma,\,p}^{\;q}=C_{\alpha\beta}K_{\gamma,\,p}^{\;q}+(\gamma_5 C)_{\alpha\beta}N_{\gamma,\,p}^{\;q}+(\mathrm{i}\gamma_\mu\gamma_5 C)_{\alpha\beta}N_{\gamma\mu,\,p}^{\;q}.$$

The contribution of $N_{[\alpha\beta]\gamma,\,p}^{\;q}$ to the currents (3·8), namely

$$J^{Ri}(N)=-\tfrac{1}{24}(\bar{N}^{[\beta\alpha]\gamma}(\gamma^R)_\alpha^{\alpha'}N_{[\alpha'\beta]\gamma})_{3D+5F}+\tfrac{1}{12}(\bar{N}^{[\beta\alpha]\gamma}(\gamma^R)_\alpha^{\alpha'}N_{[\alpha'\gamma]\beta})_{3D+2F}$$

may be now written out in terms of N and N_μ. Thus

$$J^i(N)=\tfrac{1}{2}[2\bar{N}N+\mathrm{i}(\bar{N}_\mu\gamma_\mu N-\bar{N}\gamma_\mu N_\mu)+\bar{N}_\lambda\gamma_\lambda\gamma_\mu N_\mu+\bar{N}_\mu N_\mu]^i_F,$$

$$\begin{aligned}J_\mu^i(N)=&\tfrac{1}{6}\mathrm{i}(\bar{N}_\mu N-\bar{N}N_\mu)^i_{F-3D}\\&-\tfrac{1}{6}[\mathrm{i}(\bar{N}_\lambda\gamma_\lambda\gamma_\mu N-\bar{N}\gamma_\mu\gamma_\lambda N_\lambda)-\bar{N}_\lambda\gamma_\mu N_\lambda+\bar{N}_\lambda\gamma_\lambda\gamma_\rho\gamma_\nu N_\nu]^i_{3D+2F},\end{aligned}$$

$$\begin{aligned}J_{\mu\nu}^i(N)=&\tfrac{1}{6}\mathrm{i}(\bar{N}_\mu N_\nu-\bar{N}_\nu N_\mu)^i_{F-3D}\\&+\tfrac{1}{6}[2\bar{N}\sigma_{\mu\nu}N+\mathrm{i}(\bar{N}_\lambda\gamma_\lambda\sigma_{\mu\nu}N-\bar{N}\sigma_{\mu\nu}\gamma_\lambda N_\lambda)+\bar{N}_\lambda\sigma_{\mu\nu}N_\lambda+\bar{N}_\lambda\gamma_\lambda\sigma_{\mu\nu}\gamma_\kappa N_\kappa]^i_{3D+2F},\end{aligned}$$

$$\begin{aligned}J_{\mu 5}^i(N)=&-\tfrac{1}{6}(\bar{N}\gamma_5 N_\mu+\bar{N}_\mu\gamma_5 N+\mathrm{i}\bar{N}_\lambda\gamma_\lambda\gamma_5 N_\mu-\mathrm{i}\bar{N}_\mu\gamma_5\gamma_\lambda N_\lambda)^i_{3D+5F}\\&-\tfrac{1}{6}[2(\bar{N}_\mu\gamma_5 N+\bar{N}\gamma_5 N_\mu)+(\bar{N}_\lambda\gamma_\lambda\gamma_5\gamma_\mu N+\bar{N}\gamma_\mu\gamma_5\gamma_\lambda N_\lambda)\\&\quad-2\mathrm{i}(\bar{N}_\mu\gamma_5\gamma_\lambda N_\lambda-\bar{N}_\lambda\gamma_\lambda\gamma_5 N_\mu)-\mathrm{i}\bar{N}_\lambda\gamma_\mu\gamma_5 N_\lambda+\mathrm{i}\bar{N}_\lambda\gamma_\lambda\gamma_\mu\gamma_5\gamma_\nu N_\nu]^i_{3D+2F},\end{aligned}$$

$$\begin{aligned}J_5^i(N)=&-\tfrac{1}{6}(2\bar{N}\gamma_5 N+\mathrm{i}\bar{N}_\mu\gamma_\mu\gamma_5 N-\mathrm{i}\bar{N}\gamma_5\gamma_\mu N_\mu)^i_{6D+7F}\\&+\tfrac{1}{6}(\bar{N}_\lambda\gamma_5 N_\lambda+\bar{N}_\lambda\gamma_\lambda\gamma_5\gamma_\mu N_\mu)^i_{3D+2F}.\end{aligned}$$

Note now that already the sacrosanct combinations $3D+2F$ and $3D+5F$ of $\tilde{U}(12)$ have disappeared, being replaced by their various linear combinations.

5. The inhomogeneous Lorentz group and the final expressions for the form factors

A. The work so far has been concerned with purely static considerations. We have computed the expectation values of matrices γ^R between (the homogeneous Lorentz group) multispinors of various symmetry properties. The formalism can have no physical content till these spinors are made to represent physical particles,

† $\Psi_{[\alpha\beta]\gamma}$ has 24 components; equation (4·8) states that the fully antisymmetric part of $\Psi_{[\alpha\beta]\gamma}$ vanishes. We are thus left with 20 independent components if equation (4·8) is satisfied.

i.e. till the formalism assures that they correspond to the representations of the inhomogeneous Poincaré group. One needs therefore, at this stage, some equations of motion which the spinors $\Psi_{\alpha\beta\gamma\ldots}$ must satisfy. This essentially is the point of departure of our work in comparison with other approaches to the problem and was stressed strongly in I.

Among the variety of higher spin equations available, we shall choose the simplest, and the least restrictive (though in many ways the most profound) set of equations: we generalize the Bargmann–Wigner (1948)† approach to the representations of the inhomogeneous Lorentz group. The approach works with the equations

$$\left.\begin{aligned}(\gamma p)^{\alpha'}_{\alpha}\Psi_{\alpha'\beta\gamma\ldots}(p) &= m\Psi_{\alpha\beta\gamma\ldots}(p),\\ (\gamma p)^{\beta'}_{\beta}\Psi_{\alpha\beta'\gamma\ldots}(p) &= m\Psi_{\alpha\beta\gamma\ldots}(p).\\ &\cdots\cdots\cdots\cdots\end{aligned}\right\} \tag{5·1}$$

These describe particles of (*a*) one definite mass m, (*b*) one definite spin (provided $\Psi_{\alpha\beta\gamma\ldots}$ is a spinor of a definite symmetry type), (*c*) the solutions of the equations pose no problems of negative energies or indefinite metrics and (*d*) the higher spinors transform 'visibly' as direct products of the fundamental quark.

We now examine the implications of applying these equations to the spinors of rank 2 and 3 considered previously:

(i) *Spinor of rank* 2

Write
$$\Phi^{\beta}_{\alpha} = [\phi + \gamma_5\phi_5 + i\gamma_\mu\gamma_5\phi_{\mu 5} + \gamma_\mu\phi_\mu + \tfrac{1}{2}\sigma_{\mu\nu}\phi_{\mu\nu}]^{\beta}_{\alpha}. \tag{5·2}$$

From
$$(\gamma p)^{\alpha'}_{\alpha}\Phi^{\beta}_{\alpha'} = m\Phi^{\beta}_{\alpha}, \quad -(\gamma p)^{\beta}_{\beta'}\Phi^{\beta'}_{\alpha} = m\Phi^{\beta}_{\alpha}$$

we deduce
$$\left.\begin{aligned}\phi &= 0,\\ p_\mu\phi_5 = im\phi_{\mu 5}, &\quad p_\mu\phi_{\mu 5} = -im\phi_5,\\ p_\mu\phi_\nu - p_\nu\phi_\mu = im\phi_{\mu\nu}, &\quad p_\nu\phi_{\nu\mu} = -im\phi_\mu.\end{aligned}\right\} \tag{5·3}$$

Thus $(\phi_5, \phi_{\mu 5})$ together describe (Kemmer 1939) a 0^- particle and $(\phi_\mu, \phi_{\mu\nu})$ describe a 1^- particle. The relation of the second-rank spinor with what is essentially the Kemmer theory of spin zero and spin one particles was first pointed out by Belinfante (1939; the spinor is called an undor in his terminology). Note that the assumption that the tensor $\Phi^{\beta}_{\alpha}(p)$ transforms as a quark–antiquark composite $\psi_\alpha(p)\,\overline{\psi}^{\beta}(p)$ (with no relative momentum) fixes the parities of the mesons unambiguously.‡

(ii) *Fully symmetric spinor of rank* 3

On account of full symmetry the three equations can be collapsed into a single equation
$$(\gamma p)^{\alpha'}_{\alpha}\Psi_{\alpha'\beta\gamma} = m\Psi_{\alpha\beta\gamma}.$$

Substituting the expression (4·5) into this equation and contracting it with $(C^{-1}\gamma_\mu)^{\beta\gamma}$, $(C^{-1}\gamma_\mu)^{\alpha\beta}$ and $(C^{-1}\sigma_{\mu\nu})^{\alpha\beta}$ we find

$$(\gamma p - m)\,\psi_\mu = 0, \tag{5·4}$$

$$p_\nu\psi_{\nu\mu} = -im\psi_\mu, \quad p_\mu\psi_\nu - p_\nu\psi_\mu = im\psi_{\mu\nu}. \tag{5·5}$$

† These authors state the equations for the case of full symmetry in $\alpha, \beta, \gamma, \ldots$. The extension to other irreducible spinors of mixed symmetry (defined in the manner of §4) presents no difficulties. It is this type of extension, however, which allows us to describe a spin $\frac{1}{2}$ particle, for example, by a 20-component spinor.

‡ Hereafter we assume that for the fundamental quark $P\psi P^{-1} = \gamma_0\psi$.

It is simple to show, moreover, that $\psi_\mu, \psi_{\mu\nu}$ satisfying (4·6) and (4·7) give a fully symmetrical $\Psi_{\alpha\beta\gamma}$. Thus the system is entirely equivalent to the Rarita–Schwinger formalism (Rarita & Schwinger 1941) for a particle of spin $\frac{3}{2}^+$, except that the components $\psi_{\mu\nu}$ now appear 'on par' with ψ_μ.

(iii) *Mixed spinor of rank* 3

Applied to equation (4·9), the first equation of motion

$$(\gamma p)^{\gamma'}_{\gamma}\Psi_{\alpha\beta\gamma'} = m\Psi_{\alpha\beta\gamma}$$

gives simply
$$(\gamma p - m)\,\psi_\mu = (\gamma p - m)\,\psi = (\gamma p - m)\,K = 0, \tag{5·6}$$

while the other pair of equations

$$(\gamma p)^{\alpha'}_{\alpha}\Psi_{\alpha'\beta\gamma} = m\Psi_{\alpha\beta\gamma}$$

give the relations
$$\left.\begin{aligned} K = 0, \quad p_\mu\psi_\nu - p_\nu\psi_\mu = 0, \\ p_\mu\psi = im\psi_\mu, \quad p_\mu\psi_\mu = -im\psi. \end{aligned}\right\} \tag{5·7}$$

Taken together the system clearly describes a particle of spin $\frac{1}{2}^+$.

B. With the Bargmann–Wigner equations our identification with physical particles of the $\tilde{U}(12)$ quantities Φ^A_B and Ψ_{ABC} is complete.

We here summarize the results:

(i) The regular representation $12 \times 12^*$ decomposes as

$$\Phi^A_B = (\phi^j + \gamma_5\phi^j_5 + i\gamma_\mu\gamma_5\phi^j_{\mu 5} + \gamma_\mu\phi^j_\mu + \tfrac{1}{2}\sigma_{\mu\nu}\phi^j_{\mu\nu})^\alpha_\beta\,(T^j)^p_q.$$

If the mesons are free with mass m, the implication of the equations of motion are

$$p_\mu\phi^j_5 = im\phi^j_{\mu 5}, \quad p_\mu\phi^j_{\mu 5} = -im\phi^j_5 \quad \text{for } (0^-) \text{ particles},$$

$$p_\mu\phi^j_\nu - p_\nu\phi^j_\mu = im\phi^j_{\mu\nu}, \quad p_\nu\phi^j_{\nu\mu} = -im\phi^j_\mu \quad \text{for } (1^-) \text{ particles},$$

$$\phi^j \equiv 0 \quad \text{for } (0^+).$$

Thus spin zero particles are represented by 5-component entities; spin one by ten components and

$$144 = 143 + 1 = 9 \times 10 + 9 \times 5 + 9.$$

The last nine are the so-called trivial components.

(ii) *Rank* 3

The 364 components of the fully symmetric $\tilde{U}(12)$ tensor Ψ_{ABC} decompose as

$$\underline{364} = (10, 20) + (8, 20) + (1, 4).$$

In detail:

$$\Psi_{\alpha p\beta q\gamma r} = D_{\alpha\beta\gamma, pqr} + \epsilon_{pqr}V_{[\alpha\beta\gamma]} + \frac{1}{2\sqrt{6}}(\epsilon_{pqs}N^s_{[\alpha\beta]\gamma, r} + \epsilon_{qrs}N_{[\beta\gamma]\alpha,\, p}{}^{s} + \epsilon_{rps}N_{[\gamma\alpha]\beta,\, q}{}^{s}),$$

where D is completely symmetric both in its spinor and unitary spin indices, $V_{[\alpha\beta\gamma]}$ is completely antisymmetric and $N_{[\alpha\beta]\gamma,\, r}{}^{s}$ has mixed symmetry in spinor indices and is traceless in unitary spin indices.

The equations of motion ensure that D (with is 20 components) describes a particle of spin $\frac{3}{2}$, N (with its 20 components) a particle of spin $\frac{1}{2}$ and V vanishes identically because of complete antisymmetry. The relative parities of the decimet

and the octet are the same (and for the quark ψ transforming as $P\psi P^{-1} = \gamma_0\psi$ the same as the quark). The detailed consequences of the equations are to allow us to write D and N in the forms:

$$D_{\alpha\beta\gamma,pqr}(p) = (\gamma_\mu C)_{\alpha\beta} D_{\gamma\mu,pqr}(p) + \frac{\mathrm{i}}{2m}(\sigma_{\mu\nu}C)_{\alpha\beta}[p_\mu D_{\gamma\nu,pqr}(p) - p_\nu D_{\gamma\mu,pqr}(p)],$$

with
$$(\gamma p - m) D_{\mu,pqr}(p) = 0, \quad \gamma_\mu D_{\mu,pqr}(p) = 0, \tag{5·8}$$

and for the eightfold baryon N:

$$mN_{[\alpha\beta]\gamma,r}^{s}(p) = [(\gamma p + m)\gamma_5 C]_{\alpha\beta} N_{\gamma,r}^{s}(p),$$
$$(\gamma p - m) N_r^s(p) = 0. \tag{5·9}$$

It is important to remember that the $\tilde{U}(12)$ multiplets (which were decomposed relative to $U(3)\times\mathscr{L}_4$ before the Bargmann–Wigner equations were applied) now no longer possess the full symmetry.† The exact $U(6)$ limit, however, can still be recovered, by going for every particle to its rest frame ($p = 0$) and projecting out to the positive energy subspace.‡ The demands of 'relativistic completion' are incompatible with exact symmetry at any (but zero) momenta.

One point is worth emphasizing again at this stage. With the Bargmann–Wigner equations the $\underline{364}$ multiplet of $\tilde{U}(12)$ has exactly the content of the $\underline{56}$ of $U(6)$; likewise for the $\underline{143}$ of $\tilde{U}(12)$ which corresponds with the $\underline{35}$ of $U(6)$; no more and no less. This will happen for all multiplets. In a future paper we shall treat the $\underline{572}$ multiplet and show how its algebra provides the relativistic completion of the $\underline{70}$ of $U(6)$ in a one-one manner.

C. With (5·8) and (5·9) inserted into (3·8) and p and p' denoting the incoming and outgoing baryon momenta, we can now give the final expressions for the form factors. These are

$$J^i = \tfrac{1}{4}m^{-2}P^2[(\bar{N}N)_F + 3\bar{D}_\lambda D_\lambda] + \tfrac{3}{2}m^{-2}q_\lambda \bar{D}_\lambda q_\kappa D_\kappa, \tag{5·10}$$

$$J_\mu^i = \tfrac{1}{2}m^{-1}P_\mu(\bar{N}N)_F + \tfrac{1}{4}m^{-2}(\bar{N}r_\mu N)_{D+\frac{1}{2}F} + m^{-2}(\epsilon_{\mu\nu\kappa\lambda}P_\kappa q_\lambda \bar{D}_\nu N + \text{h.c.})$$
$$+ \tfrac{3}{4}m^{-2}P^2\bar{D}_\lambda\gamma_\mu D_\lambda + \tfrac{3}{2}m^{-2}q_\lambda\bar{D}_\lambda\gamma_\mu q_\kappa D_\kappa, \tag{5·11}$$

$$J_{\mu\nu}^i = \tfrac{1}{4}m^{-2}P^2[(\bar{N}\sigma_{\mu\nu}N)_{D+\frac{1}{2}F} + 3\bar{D}_\lambda\sigma_{\mu\nu}D_\lambda] + \tfrac{3}{2}m^{-2}q_\lambda\bar{D}_\lambda\gamma_\mu q_\kappa D_\kappa$$
$$+ \tfrac{1}{4}\mathrm{i}m^{-2}(P_\mu q_\nu - P_\nu q_\mu)(\bar{N}N)_{\frac{1}{2}F-D} + \mathrm{i}m^{-1}(\epsilon_{\mu\nu\kappa\lambda}P_\lambda\bar{D}_\kappa N + \text{h.c.}), \tag{5·12}$$

$$J_{\mu 5}^i = \tfrac{1}{4}m^{-2}P^2(\mathrm{i}\bar{N}\gamma_\mu\gamma_5 N)_{D+\frac{1}{2}F} + \tfrac{3}{4}\mathrm{i}m^{-2}P^2\bar{D}_\lambda\gamma_\mu\gamma_5 D_\lambda + \tfrac{3}{2}\mathrm{i}m^{-2}q_\lambda\bar{D}_\lambda\gamma_\mu\gamma_5 q_\nu D_\nu$$
$$- \tfrac{1}{2}\mathrm{i}m^{-2}P^2(\bar{D}_\mu N + \text{h.c.}) + \mathrm{i}m^{-2}(p'_\mu p_\lambda \bar{D}_\lambda N + \text{h.c.}), \tag{5·13}$$

† The residual symmetry will always depend on the type of equation applied to the spinor $\Psi_{\alpha\beta\gamma\ldots}$. For example if in place of equations (5·1), one had used equation (8) of paper I ($T\{\gamma P + \mathrm{i}\gamma^5\gamma W\}\psi = M\psi$) the $W(6)$ symmetry would have formally survived.

‡ We shall here prove that the positive energy projection of all irreducible representations $\Psi_{\alpha\beta\gamma\ldots}(p)$ of $\tilde{U}(12)$ remain irreducible representations of $U(6)$ even after Bargmann–Wigner equations are applied provided $\mathbf{p} = 0$. Here by $U(6)$ we mean the group generated by the matrices T^i, $T^i\sigma$ (σ are 2×2 Pauli matrices). The proof is elementary. For $\mathbf{p} = 0$ the positive energy projection of γp equals $m\mathbf{1}$, and $\alpha, \beta, \gamma, \ldots$ run over 1, 2 rather than 1, ..., 4. Thus the symmetry character of any Ψ is unaltered by the application of the set of equations and carries itself from $\tilde{U}(12)$ to $U(6)$.

$$J_5^i = \tfrac{1}{4}m^{-2}P^2(\bar{N}\gamma_5 N)_{D+\frac{2}{3}F} + m^{-1}(q_\lambda \bar{D}_\lambda N + \text{h.c.}) + \tfrac{3}{4}m^{-2}P^2\bar{D}_\lambda\gamma_5 D_\lambda + \tfrac{3}{2}m^{-2}q_\lambda \bar{D}_\lambda \gamma_5 q_\kappa D_\kappa, \quad (5{\cdot}14)$$

where $P = p+p'$, $q = p-p'$; $r_\mu \equiv \epsilon_{\mu\nu\kappa\lambda}P_\nu q_\kappa \gamma_\lambda\gamma_5$. Note that the coefficient multiplying $P_\mu/2m$ and $r_\mu/4m^2$ are the Sachs (1962) form factors F_C and F_M respectively. To recover the conventional $U(6)$ results for the form factors (Sakita 1964; Bég, Lee & Pais 1964) take $\mathbf{p} = \mathbf{p}' = 0$.

If we further make the identifications

$$i\mu\phi^i_{\kappa\lambda} = p_\kappa\phi^i_\lambda - p_\lambda\phi^i_\kappa,$$

$$i\mu\phi^i_{\lambda 5} = p_\lambda\phi^i_5, \quad \mu = \text{meson mass},$$

for the meson field interaction $\Phi^B_A J^A_B$, we arrive at the following predictions for the pseudoscalar and vector currents:

$$J_5 = \left(1+\frac{2m}{\mu}\right)\frac{P^2}{4m^2}[(\bar{N}\gamma_5 N)_{D+\frac{2}{3}F} + 3\bar{D}_\lambda\gamma_5 D_\lambda] + \frac{3}{2m^2}\left(1+\frac{2m}{\mu}\right)q_\lambda\bar{D}_\lambda\gamma_5 q_\kappa D_\kappa + \frac{1}{m}\left(1+\frac{2m}{\mu}\right)[q_\lambda\bar{D}_\lambda N + \text{h.c.}], \quad (5{\cdot}15)$$

$$J_\mu = \frac{P_\mu}{2m}\left(1+\frac{q^2}{2\mu m}\right)(\bar{N}N)_F + \left(1+\frac{2m}{\mu}\right)\left(\bar{N}\frac{r_\mu}{4m^2}N\right)_{D+\frac{2}{3}F} + \frac{3P^2}{4m^2}\bar{D}_\lambda\left[\left(1+\frac{2m}{\mu}\right)\gamma_\mu - \frac{P_\mu}{\mu}\right]D_\lambda + \frac{3}{2m^2}q_\lambda\bar{D}_\lambda\left[\left(1+\frac{2m}{\mu}\right)\gamma_\mu - \frac{P_\mu}{\mu}\right]q_\kappa D_\kappa - \frac{1}{m^2}\left(1+\frac{2m}{\mu}\right)[\epsilon_{\mu\nu\kappa\lambda}P_\nu q_\kappa\bar{D}_\lambda N + \text{h.c.}]. \quad (5{\cdot}16)$$

6. Results and conclusions

As shown in § 5C the baryon–meson (N+NM) vertex contains just one form factor. Its relation with the electric and the magnetic form factors introduced by Sachs (1962) is obvious. Assuming that the photon ($F_{\mu\nu}$) couples with the composite structure of the baryon† through an effective gauge-invariant 'interaction' $f(q^2)F_{\mu\nu}M^u_{\mu\nu}$, where $M^u_{\mu\nu}$ is the U-scalar meson combination ($M^u = M^3 + (3)^{-\frac{1}{2}}M^8$), one would get for the electromagnetic form factors the expressions

$$J^C_\mu = \frac{P_\mu}{2m}\left(1+\frac{q^2}{2\mu m}\right)(\bar{N}N)_F F(q^2), \quad (6{\cdot}1)$$

$$J^M_\mu = \left(1+\frac{2m}{\mu}\right)\left(\bar{N}\frac{r_\mu}{4m^2}N\right)_{D+\frac{2}{3}F}, \quad (6{\cdot}2)$$

† To write a minimal photon interaction ($\gamma p \to \gamma p - e\gamma A$) seems highly inappropriate for a structure as complex as a baryon. To take an analogy, no one would normally contemplate using the 'minimal ansatz' for a helium 3 or tritium nucleus.

where $F(q^2) \propto f(q^2)/(q^2-\mu^2)$. As stated in the introduction we note that

(i)
$$\left.\begin{aligned}\mu_p &= (1+2m/\mu) \quad \text{(Bohr magnetons)},\\ \mu_n &= -\tfrac{2}{3}(1+2m/\mu).\end{aligned}\right\} \tag{6·3}$$

(ii)
$$\frac{J^C}{J^M} = \frac{1+q^2/2\mu m}{1+2m/\mu}. \tag{6·4}$$

For small q^2 essentially there is therefore just one electromagnetic (Sachs) form factor (Barnes 1962).

(iii) The U_6 limit of $\bar{U}(12)$ taken at the stage of equations (5·10) to (5·14) results in losing the 'anomalous' magnetic moment of the proton. Stated otherwise, the extension of the ($U(6)$) 35-fold of mesons to a ($\bar{U}(12)$) 143-fold is essential to obtain the results (6·3).

(iv) For weak interactions (5·11) and (5·13) give the vector and axial-vector form factors. These reproduce the well-known $U(6)$ result $g_A/g_v = -5/3$ at zero momentum transfer (Kawarabayashi 1964; Rosen & Parksava 1964; Bég & Pais 1965; Altarelli, Buccella & Gatto 1964; Babu 1964).

(v) *Meson–meson vertex*

In general $\underline{143} \times \underline{143}$ contains $\underline{143}$ twice but for the special case of identical $\underline{143}$'s, there is just one coupling of the type

$$\mathscr{L} = \Phi^B_A \Phi^C_B \Phi^A_C.$$

Written out in terms of ϕ^i_5, ϕ^i_μ, etc., introduced in equation (5·2), $\mathscr{L}$ equals

$$f^{ijk}[3\phi^i_\mu \phi^j_\nu \phi^k_{\mu\nu} + 3\phi^i_{\mu 5}\phi^j_{\mu\nu}\phi^k_{\nu 5} + 6\phi^i_5 \phi^j_\mu \phi^k_{\mu 5} + \phi^i_{\mu\nu}\phi^j_{\nu\lambda}\phi^k_{\lambda\mu}]$$
$$+ \tfrac{3}{2} d^{ijk}\epsilon_{\kappa\lambda\mu\nu}[2\phi^i_k \phi^j_{\lambda\mu}\phi^k_{\nu 5} + \tfrac{1}{2}\phi^i_{\kappa\lambda}\phi^j_{\mu\nu}\phi^k_5 + (\text{terms in } \phi^i)]. \tag{6·5}$$

Introducing equations (5·3), we get for the effective vector current

$$J^i_\mu(p,p') = f^{ijk}\Big[-P_\mu\Big\{\Big(1+\frac{q^2}{2\mu^2}\Big)\phi^j_\nu(p')\,\phi^k_\nu(-p) - \Big(1+\frac{q^2}{2\mu^2}\Big)\phi^j_5(p')\,\phi^k_5(-p)\Big\}$$
$$+ 3\{q_\nu \phi^j_\nu(p')\,\phi^k_\mu(-p) + q_\nu \phi^j_\nu(-p)\,\phi^k_\mu(p')\} - P_\mu q_\nu \phi^j_\nu(p')\,q_\lambda \phi^k_\lambda(-p)/\mu^2\Big]$$
$$-\frac{i}{\mu} d^{ijk}\epsilon_{\kappa\lambda\mu\nu}P_\lambda q_\nu\{\phi^j_k(p')\,\phi^k_5(-p) + \phi^j_5(p')\,\phi^k_\kappa(-p)\}. \tag{6·6}$$

Introducing the gauge-invariant electromagnetic coupling $f(q^2)F_{\mu\nu}\phi_{\mu\nu}$, we obtain a total magnetic moment of 3 Bohr magnetons and a quadrupole moment of -4 (in units of e/μ^2) for positively charged 1^- particles.

(vi) *The outlook*

With the effective baryon–meson and meson–meson vertices available it is a trivial step to write pole approximations for the strong interaction four-particle process. With this approximation as the starting-point all S-matrix techniques

(like Mandelstam representation, Reggeization, analytic continuation now both in angular momentum and unitary spin) are available for determining the complete ($\bar{U}(12)$) relativistic S-matrix theory. This is so because as a rule all that the S-matrix theory requires are 'Born approximations' as the 'input'. The higher 'fundamental' 4-particle vertices involving 4-particles (like $\bar{N}NMM$) which are analogous to 4-field Lagrangians (and knowledge of which is necessary for peripheral interactions) are no harder to write down, using the methods and techniques of this paper. Every one of these 'fundamental' interactions will now be momentum dependent—this dependence coming in a determinate manner from $\bar{U}(12)$ symmetry. The conventional field theory Lagrangians are but local approximations of these 'fundamental' interactions.

Note that one of the ways in which the inhomogeneous Lorentz group breaks the full $\bar{U}(12)$ symmetry is to force certain components of Φ_B^A to become 'redundant'. In a future paper we wish to exploit these 'redundant' components and use them as symmetry-breaking 'spurions'.

It would seem that the $\bar{U}(12)$ ideas fulfil completely the dream of the $U(6)$ theorist, i.e. to write S-matrix elements whose momentum dependence is dictated by the internal symmetry group of SU_3.

Our thanks are due to Dr M. A. Rashid for helping with the content of the $\bar{U}(12)$ representations. We are grateful to Drs J. Charap, P. T. Matthews, and J. C. Ward for numerous fruitful discussions.

Appendix

Our convention for the unitary spin matrices is the following: $T^i = \frac{1}{2}\lambda^i$, with the λ^i defined by Gell-Mann (1962). Thus $\mathrm{Tr}(T^iT^j) = \frac{1}{2}\delta^{ij}$, $f^{0jk} = 0$, and $d^{0jk} = \delta^{jk}(2/3)^{\frac{1}{2}}$.

$$[T^i, T^j] = \mathrm{i}f^{ijk}T^k, \quad \{T^i, T^j\} = d^{ijk}T^k.$$

From the fundamental representation, $F^{Ri} \equiv \gamma^R T^i$ we deduce the following commutators:

$$[F^i, F^j] = \mathrm{i}f^{ijk}F^k,$$

$$[F^i, F_5^j] = \mathrm{i}f^{ijk}F_5^k,$$

$$[F_5^i, F_5^j] = -\mathrm{i}f^{ijk}F^k,$$

$$[F^i, F_{\mu\nu}^j] = \mathrm{i}f^{ijk}F_{\mu\nu}^k,$$

$$[F_5^i, F_{\mu\nu}^j] = \tfrac{1}{2}\mathrm{i}f^{ijk}\epsilon_{\mu\nu\kappa\lambda}F_{\kappa\lambda}^k,$$

$$[F_{\kappa\lambda}^i, F_{\mu\nu}^j] = \mathrm{i}d^{ijk}(g_{\kappa\nu}F_{\lambda\mu}^k + g_{\lambda\mu}F_{\kappa\nu}^k - g_{\kappa\mu}F_{\lambda\nu}^k - g_{\lambda\nu}F_{\kappa\nu}^k)$$
$$+ \mathrm{i}f^{ijk}\{(g_{\kappa\mu}g_{\lambda\nu} - g_{\lambda\mu}g_{\kappa\nu})F^k - \epsilon_{\kappa\lambda\mu\nu}F_5^k\},$$

$$[F_\mu^i, F_\nu^j] = \mathrm{i}f^{ijk}g_{\mu\nu}F^k - \mathrm{i}d^{ijk}F_{\mu\nu}^k,$$

$$[F_\mu^i, F_{\nu 5}^j] = \mathrm{i}d^{ijk}g_{\mu\nu}F_5^k + \tfrac{1}{2}\mathrm{i}f^{ijk}\epsilon_{\mu\nu\kappa\lambda}F_{\kappa\lambda}^k,$$

$$[F_{\mu 5}^i, F_{\nu 5}^j] = -\mathrm{i}f^{ijk}g_{\mu\nu}F^k + \mathrm{i}d^{ijk}F_{\mu\nu}^k,$$

and

$$[F^i, F^j_\mu] = \mathrm{i}f^{ijk}F^k_\mu,$$

$$[F^i, F^j_{\mu 5}] = \mathrm{i}f^{ijk}F^k_{\mu 5},$$

$$[F^i_5, F^j_\mu] = \mathrm{i}d^{ijk}F^k_{\mu 5},$$

$$[F^i_5, F^j_{\mu 5}] = \mathrm{i}d^{ijk}F^k_\mu,$$

$$[F^i_\lambda, F^j_{\mu\nu}] = \mathrm{i}d^{ijk}(g_{\lambda\mu}F^k_\nu - g_{\lambda\nu}F^k_\mu) - \mathrm{i}f^{ijk}\epsilon_{\lambda\mu\nu\kappa}F^k_{\kappa 5},$$

$$[F^i_{\lambda 5}, F^j_{\mu\nu}] = \mathrm{i}d^{ijk}(g_{\lambda\mu}F^k_{\nu 5} - g_{\lambda\nu}F^k_{\mu 5}) - \mathrm{i}f^{ijk}\epsilon_{\lambda\mu\nu\kappa}F^k_\kappa.$$

References

References A

Salam, A. 1965 *Phys. Lett.* (in the Press).
Delbourgo, R., Salam, A. & Strathdee, J. 1964*a* ICTP 64/7; 1964*b* ICTP 64/11 (referred to as I) (to be published).
Marshak, R. E. & Okubo, S. 1964 *Phys. Rev. Lett.* **13**, 818, and preprint (Rochester).
Feynmann, R. P., Gell-Mann, M. & Zweig, G. 1964 *Phys. Rev. Lett.* **13**, 678.
Bardacki, K., Cornwall, J. M., Freund, P. G. O. & Lee, B. W. 1964, 1965 *Phys. Rev. Lett.* **13**, 698; **14**, 48.
Fulton, T. & Wess, J. 1964 Preprints (Vienna).
Agasshi, J. J. & Roman, P. 1964 Preprint (Boston).
Bég, M. A. B. & Pais, A. 1965 *Phys. Rev.* (to be published).
Mahanthrappa, K. T. & Sudarshan, E. C. G. 1965 Preprint (Syracuse).
Wyld, H. W. 1965 Preprint (Illinois).

Other references

Altarelli, G., Buccella, F. & Gatto, R. 1964 Preprints (Florence).
Babu, P. 1964 Preprint (Bombay).
Bargmann, V. & Wigner, E. 1948 *Proc. Nat. Acad. Sci., Wash.*, **34**, 211.
Barnes, K. J. 1962 *Phys. Rev. Lett.* **1**, 166.
Bég, M. A. B., Lee, B. W. & Pais, A. 1964 *Phys. Rev. Lett.* **13**, 514.
Bég, M. A. B. & Pais, A. 1965 *Phys. Rev. Lett.* **14**, 51.
Belinfante, F. J. 1939 *Physica*, **6**, 870.
Gell-Mann, M. 1962 *Phys. Rev.* **125**, 1067.
Gell-Mann, M. 1964 *Phys. Lett.* **8**, 216.
Jauch, J. M. & Rohrlich, F. 1955 *The theory of photons and electrons*. New York: Addison-Wesley.
Kawarabayashi, K. 1964 ICTP 64/5.
Kemmer, N. 1939 *Proc. Roy. Soc.* A, **173**, 91.
Rarita, W. & Schwinger, J. 1941 *Phys. Rev.* **60**, 61.
Rosen, S. P. & Paksava, S. 1964 *Phys. Rev. Lett.* **13**, 773.
Sachs, R. G. 1962 *Phys. Rev.* **126**, 2256.
Sakita, B. 1964 *Phys. Rev. Lett.* **13**, 643.
Zweig, G. 1964 C.E.R.N. Rep. (unpublished).

Volume 25B, number 3 PHYSICS LETTERS 21 August 1967

HARMONIC ANALYSIS IN TERMS OF THE HOMOGENEOUS LORENTZ GROUP*

R. DELBOURGO
Imperial College, London

A. SALAM** and J. STRATHDEE
International Centre for Theoretical Physics, Trieste

Received 7 June 1967

Scattering amplitudes for particles of arbitrary mass and spin are expanded in terms of homogeneous Lorentz group representations. The expansion satisfies the group theoretic O(3,1) boundary constraints in the forward elastic limit and incorporates the correct threshold factors.

Following Toller's realization [1] that the elastic forward scattering amplitude admits of an O(3,1) little group invariance on account of vanishing momentum transfer four-vector, considerable interest has lately arisen in partial wave analyses using unitary representations of the homogeneous Lorentz group [2].

The physical consequences of using this type of expansion are two-fold; first the solution of $t = 0$ singularity problems [3] which had made t .r appearance in Regge analyses based on the standard angular momentum O(3) expansion (evasions versus conspiracies); second, the appearance of "daughter" poles associated with a (Toller) pole in the complex O(3,1) plane.

This note is concerned with two important generalizations:

(A) Expansion of the flipless elastic amplitude for all values of momentum transfer and

(B) Extension of the expansion to the spin-flip amplitudes.

Our methods only use group theory to provide the boundary conditions on the suggested expansions; otherwise they are simply based upon the fact that any function $T(\zeta)$ is square integrable*** in the sense that

$$\int_0^\infty d\zeta \, \mathrm{sh}^2 \zeta \, |T(\zeta)|^2 < \infty ,$$

can be expanded in terms of the principal series of functions of the homogeneous Lorentz group. Specifically,

$$T_{S'\lambda S}(\zeta) = \sum_{|j_0| \geq \min(S,S')} \int_0^{i\infty} i \, d\sigma (j_0^2 - \sigma^2) \times \times T_{S'S}(j_0, \sigma) \, d^{j_0 \sigma^*}_{S\lambda S}(\zeta) . \quad (1)$$

Here σ and j_0 are related to the two Casimir operators of O(3,1); $0 < \sigma < i\infty$ and j_0 is discrete. Thus consider the elastic scattering amplitude in the centre-of-mass and define a reduced amplitude $T_{S'\lambda'S\lambda}$ by

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research OAR through the European Office Aerospace Research, United States Air Force.

** On leave of absence from Imperial College, London.

*** The (near) forward scattering amplitudes do not satisfy the square-integrability criterion. However, one may assume that for large negative t, there exist a region of s where the criterion is satisfied. The expansion can then be continued in t to small negative t and eventually to $t > 0$. In this continuation process one will pick up poles in the complex σ-plane which pierce the background integral. It is possible that cut singularities could be associated with the supplementary series of the Lorentz group; thus if $T(\zeta)$ is not square-integrable, the expansion (1) may need also a supplementary series integral ranging from $0 < \sigma < 1$, and it appears tempting to postulate that the existence of such an integral can be associated with the cut phenomena in the complex σ-plane.

Reprinted with permission from *Phys. Lett.* **25B** (1967) pp. 230–232
© 1967 Elsevier Science Publishers B.V.

$$\langle p_3 S_3 \lambda_3, p_4 S_4 \lambda_4 | T | p_1 S_1 \lambda_1, p_2 S_2 \lambda_2 \rangle =$$

$$= \sum_{S\lambda, S'\lambda'} \langle S_4 \lambda_4 | S_2 \lambda_2, S' - \lambda' \rangle \times$$

$$\times T_{S'\lambda', S\lambda}(s,t) \langle S_3 \lambda_3, S\lambda | S_1 \lambda_1 \rangle \quad (2)$$

Here the amplitude on the left is the conventional helicity amplitude with λ's as the eigenvalues of J_{12} in the respective rest frames.

(A) In the forward direction $T_{S'\lambda' S\lambda}(s, t=0) = \delta_{\lambda\lambda} T_{S'\lambda S}(s)$, has the group theoretic expansion (1) (with $\operatorname{ch}\zeta = (s-u)/4m_1 m_2$), and $\lambda = \lambda'$. For $t \neq 0$ ($\lambda = \lambda'$), we use eq. (1) to write

$$T_{S'\lambda, S\lambda}(s,t) = \sum_{j_0} \int i\, d\sigma\, (j_0^2 - \sigma^2) \times$$

$$\times T_{S'S}(j_0, \sigma, t)\, d^{j_0 \sigma^*}_{S\lambda S'}(\zeta_t) \quad (3)$$

with

$$\operatorname{ch}\zeta_t = (\widehat{p_1 + p_3}) \cdot (\widehat{p_2 + p_4}) =$$

$$= (s-u)[(4m^2 - t)(4m'^2 - t)]^{-\frac{1}{2}} \quad (4)$$

Clearly eq. (3) satisfies all group-theoretic boundary conditions at $t=0$.

(B) In order to see how to incorporate the spin-flip amplitudes, which reduce to eq. (3) when there is no spin-flip, note from the standard O(3) partial wave expansion

$$T_{(\lambda)}(s,t) = \sum_j (2j+1)\, T^j_{(\lambda)}(s)\, d^j_{\lambda_1 - \lambda_2,\, \lambda_3 - \lambda_4}(\theta_s) \quad (5)$$

that for small t, we have the threshold behavior * $(\sin\tfrac{1}{2}\theta_s)^{|\Delta|} \sim t^{\frac{1}{2}|\Delta|}$ where

$$\Delta = (\lambda_1 - \lambda_2) - (\lambda_3 - \lambda_4) = (\lambda_1 - \lambda_3) - (\lambda_2 - \lambda_4). \quad (6)$$

Thus we should anticipate the occurrence of spin-flip factors $t^{\frac{1}{2}|\Delta}$ in the amplitude. This is in fact borne out by the group theory which necessitates that in an expansion in powers of momentum transfer,

$$T_{S'\lambda', S\lambda}(s,t) = \sum_n t^n \langle S'\lambda' | \Gamma^n_\Delta(s,t) | S\lambda \rangle, \quad (7)$$

* In general $\sin^2\theta_s = 4s\phi(s,t)/\Delta(s, m_1^2, m_2^2)\,\Delta(s, m_3^2, m_4^2)$ where $\phi(s,t) = 0$ denotes the boundary curve of the physical region and the Δ have the usual significance in terms of the relative momenta in the centre of mass frame. For equal masses, $m_1 = m_3 = m$, $m_2 = m_4 = m'$, $\sin^2\theta_s = 4st\{su - (m^2 - m'^2)^2\}/\{s-(m-m')^2\}\{s-(m+m')^2\}$.

the factor $t^{\frac{1}{2}|\Delta|}$ must be associated with the (rotation) tensor operator $\Gamma^n_\Delta(s,t)$ of order n - finite near $t=0$ - with $n \geq |\Delta|$. In general we can extract spin-flip factors from eq. (7) in (at least) two ways †

$$T_{S'\lambda', S\lambda}(s,t) = \sum_{nJ'} t^n \langle S'\lambda' | \Gamma^n_\Delta(0,0) | J'\lambda \rangle \times$$

$$\times \langle S'J'\lambda | T(s,t) | S\lambda \rangle = \quad (7')$$

$$= \sum_{J', n \geq |\Delta|} t^n (S' \| \Gamma^n \| J') \langle S'\lambda' | n\Delta, J'\lambda \rangle \langle S'J' | T_\lambda | S \rangle$$

or

$$T_{S'\lambda', S\lambda}(s,t) = \sum_{nJ} t^n \langle S'\lambda' | T(s,t) | SJ\lambda \rangle \langle J\lambda' | \Gamma^n_\Delta | S\lambda \rangle =$$

$$= \sum_{J, n \geq |\Delta|} t^n \langle S' | T_{\lambda'} | SJ \rangle (J \| \Gamma^n \| S) \langle J\lambda' | n\Delta, S\lambda \rangle. \quad (7'')$$

We now have only to restrict the variation of n and J' (or J) in order to obtain a linearly independent set of reduced amplitudes, $(2S+1)(2S'+1)$ in number. The simplest variation we have found is to set $n = |\Delta|$ and normalize $(S' | \Gamma^n | J') = 1$. The amplitudes $T_{S'\lambda', J\lambda'}$ or $T_{S'\lambda, J\lambda}$ being flipless are Tollerizable and we can therefore write

$$T_{S'\lambda', S\lambda}(s,t) =$$

$$= \sum_{J'} t^{\frac{1}{2}|\Delta|} \langle S'\lambda' | \Delta | \Delta, J'\lambda \rangle \langle S'J' | T_\lambda | S \rangle, \quad (8)$$

or

$$T_{S'\lambda', S\lambda}(s,t) =$$

$$= \sum_J t^{\frac{1}{2}|\Delta|} \langle S' | T_{\lambda'} | SJ \rangle \langle J\lambda' | \Delta | \Delta, S\lambda \rangle. \quad (8')$$

One may get an undercount of the amplitudes through the following accident; if λ (or λ') vanish with $\lambda' \neq 0$ (or $\lambda \neq 0$) using eq. (8) we would get the expansions

$$T_{S'\lambda', S0} = \sum_{J'} t^{|\Delta|} \langle S'\lambda' | \lambda' | \lambda', J'0 \rangle \langle S'J' | T_0 | S \rangle$$

$$T_{S'-\lambda', S0} = \sum_{J'} t^{|\Delta|} \langle S'\lambda' | \lambda', J'_0 \rangle \times$$

$$\times (-)^{|\lambda'| + J' - S'} (S'J' | T_0 | S).$$

† For details see Delbourgo et al. [4].

‡ Up to the imposition of discrete C, P, T symmetries.

Thus for even values of $|\lambda'| + J' - S'$ the amplitudes $T_{S'\lambda',S0}$ and $T_{S'-\lambda',S0}$ are no longer independent. This difficulty is simply circumvented if one considers the four sectors (i) $\lambda \neq 0$, $\lambda' \neq 0$, (ii) $\lambda = 0$, $\lambda' = 0$, (iii) $\lambda = 0$, $\lambda' \neq 0$, (iv) $\lambda \neq 0$, $\lambda' = 0$ separately. For sector (iii) the appropriate expansion is eq. (8'); for (iv) it is eq. (8) while for sectors (i) and (ii) it is immaterial if we expand using eqs. (8) or (8').

With eqs. (8) or (8') we are now able to make the desired expansion for the new reduced amplitudes,

$$\langle S'J'|T_\lambda|S\rangle =$$

$$= \sum_{j_0} \int i\,d\sigma (j_0^2 - \sigma^2)\, T^{(S')}_{J'S}(j_0, \sigma, t)\, d^{j_0\sigma *}_{S\lambda J'}(\zeta_t). \quad (9)$$

One may now approximate to these amplitudes by a set of Toller poles. The important point is that in a scheme of Tollerization, the high-energy behaviour of the amplitude which depends on an asymptotic expansion of the functions $d^{j_0\sigma}_{J\lambda S}(\zeta)$ does not depend on the quantum numbers J and S but only on σ, j_0 and λ;

$$(d^{j_0}(\zeta) \sim (\mathrm{ch}\,\zeta)^{\sigma - 1 - |j_0 - \lambda|}).$$

The physical consequences of introducing Toller poles into eq. (8) at $\sigma - 1 = a(j_0, t)$ are to provide a set of integrally spaced Regge daughter trajectories (n) at $\alpha_n(t) = a(j_0, t) - n$, parallel to the parent, with their residues related to the leading one simply through factors which enter the group-theoretic decomposition of O(3,1) into O(2,1) representation. This has been discussed in detail by Toller and Sciarrino [5] and one can take over their formalism without change. As yet the experimental evidence for daughters is scanty; however, if daughter trajectories do exist and, further, if the trajectories are experimentally found parallel or near parallel then our formalism will prove useful.

The unequal mass case the previous arguments require little modification. Thus we propose the very same expansion (2), (8), (9) with the same definition of ζ_t except that we replace $t^{\frac{1}{2}|\Delta|}$ by the threshold factor $|\phi(s,t)|^{\Delta}$ where $\phi(s,t) = 0$ is the curve which denotes the boundary of the physical region. The group-theoretic necessity for this change can be seen as follows. The formalism we have presented is based on a completeness relation for the functions $d^{j_0\sigma}_{S\lambda S'}$ and the possibility of expansion of a square integrable function in terms of these. We had to guarantee at each stage that the group theoretic boundary conditions for the equal mass case at $t = 0$ or more exactly at $\theta_s = 0$ were respected. For unequal mass $\phi(s,t) \sim (\sin^2\theta_s)$. The new point that emerges is the occurrence of singular residues (corresponding to factors $t^{-1}(m_1^1 - m_3^2)$ and $t^{-1}(m_2^2 - m_4^2)$) if one makes a Regge daughter decomposition * and passes from an expansion based on $d^{j_0\sigma}(\zeta_t)$ to the normal expansion $d^j(\theta_t)$.

The formalism presented in this note is as simple as the conventional O(3) expansion and we believe that the decomposition in terms of O(3,1) representations will replace the normal O(3) analysis as far as Reggeization is concerned. We strongly suggest that the experimental data be analysed afresh from this point of view **.

References

1. M. Toller, CERN Report 67/582/5 - Th. 780. This contains an extensive list of past references and summarises the author's own work on the topic; J. F. Boyce, R. Delbourgo, A. Salam and J. Strathdee, Trieste preprint IC/67/9.
2. R. Oakes, Phys. Letters 24B (1967) 154; G. Domokos, Phys. Rev., to be published; D. Freedman and J. Wang, Phys. Rev. Letters 18 (1967) 863. Actually these references use the continuation of the amplitude to the unphysical Euclidean region where the relevant little group is the compact O(4).
3. R. Volkov and V. N. Gribov, Soviet Phys., JETP 17 (1963) 720; M. L. Goldberger and C. Jones, Phys. Rev. 150 (1966) 1260; D. Freedman and J. Wang, Phys. Rev. 153 (1967) 1596; G. Domokos and P. Suranyi, Nucl. Phys. 54 (1964) 529; L. Durand, Phys. Rev. 154 (1967) 1537; J. C. Taylor, Oxford preprint 19-67.
4. R. Delbourgo, A. Salam and J. Strathdee, ICTP, Trieste, preprint IC/67/21.
5. A. Sciarrino and M. Toller, Internal Report University of Rome, No. 103 (1966).

* To make this decomposition one requires the generelisation and continuation of the usual addition theorem,

$$C_n^1(\mathrm{ch}\,\zeta_t) = \sum_k R_{kn}\, C_k^{n+1-k}(\mathrm{ch}\,\psi)\, C_k^{n+1-k}(\mathrm{ch}\psi')\, P_{nk}(\cos\theta_t),$$

with $\mathrm{ch}\zeta_t = \mathrm{ch}\psi\,\mathrm{ch}\psi' - \mathrm{sh}\psi\,\mathrm{sh}\psi'\cos\theta_t$, and $\mathrm{ch}\psi = (m_1^2 - m_3^2)\{t(t - 2m_1^2 - 2m_3^2)\}^{-\frac{1}{2}}$ $\mathrm{ch}\psi' = (m_2^2 - m_4^2) \times \{t(t - 2m_2^2 - 2m_4^2)\}^{-\frac{1}{2}}$ corresponding to the singular contributions.

** For these new analyses use must be made of the $d(\zeta_t)$ functions (which are listed in ref. 1) and their analytic continuation.

Reprinted from THE PHYSICAL REVIEW, Vol. 172, No. 5, 1727–1736, 25 August 1968

Reggeization of Quark Number

R. DELBOURGO
Imperial College, London, England

AND

ABDUS SALAM* AND J. STRATHDEE
International Atomic Energy Agency, International Centre for Theoretical Physics, Trieste, Italy

(Received 13 March 1968)

In earlier work the Reggeization of approximate dynamical groups has been presented as a calculational method for classifying particles and evaluating S-matrix elements at high energies. In continuation of this work, an especially simple model is considered where just one invariant of the higher $U(6)\times U(6)$ approximate symmetry, quark-plus-antiquark number, is Reggeized. The resulting classification of particles (according to their quark content) into exploding supermultiplets of spin and unitary spin, and the formulas for computing S-matrix elements, are given for high energies, where an exchange of an N-plane trajectory in the cross channel may be expected to dominate the scattering. The hope is that this analysis may help reduce the large number of parameters now used in Regge theory by combining Regge ideas with higher symmetries. The type of Fourier expansion on a higher approximate symmetry group and the Regge technique used here for evaluating asymptotic behavior may possess wider applications than the case considered in this paper.

1. REGGE MODEL OF HIGHER SYMMETRIES

THE Regge method in strong-interaction physics originated in the study of the S matrix for complex values of angular momentum and has recently met with a certain number of successes in describing elastic and inelastic two-body reactions. Even where it has succeeded, however, it has been necessary to admit a large number of residue parameters with no guiding principle to limit their arbitrariness. A similar situation prevailed in the absorption-model description of low-energy scattering; recently, however, higher supermultiplet theories [and in particular $U(6,6)$] were used with fair success to constrain strongly the values of the coupling-constant parameters[1] that entered into the Born approximation. One may expect that a marriage of supermultiplet schemes with Regge theory would be desirable in that it may suitably reduce the number of Regge parameters. We shall describe below one attempt[2] at obtaining these correlations based upon a supermultiplet scheme that Reggeizes the quark number.

The basis of our scheme is the following. Angular momentum is but one of the conserved quantities on which the S matrix depends. In particular, if a system possesses a higher spin-containing symmetry, there may be other conserved quantities (Casimir invariants of the relevant symmetry group) which it may be more profitable to continue to complex values and Reggeize. For example, with the hydrogen atom it is well known that one obtains a deeper insight into the dynamics of the bound states if it is the principal quantum number [connected with the well-known $O(4)$ symmetry of the system] that is Reggeized rather than the angular momentum. For hadron physics the $U(6)\otimes U(6)$ group appears to be an *approximate* symmetry for classification of particles. The analogy of the principal quantum number for the hydrogen case here would seem to be with the total quark number N (half the number of quarks plus antiquarks) and an analogous Reggeization of this number appears to be indicated. One may now go further and explore the dynamical consequences for high-energy scattering of such a Reggeization procedure and it is this aspect of the scheme in terms of its practical applications which we wish to stress in this paper.[3]

The consequences of the scheme are twofold:

(i) One obtains two master trajectories [plots of ReN versus (mass)^2], one for mesons ($B=0$) and one for fermions ($B=1$). For $M^2>0$, ReN goes through 1, 2, 3, $\cdots$ for mesons and $\frac{3}{2}$, $\frac{5}{2}$, $\cdots$ for fermions. On present evidence it is not excluded that this simple picture of Regge recurrences classified according to quark content can accommodate all known semistable meson and baryon states. The idea that there should be basically only one baryon and one meson entity was proposed long ago by Weisskopf.

(ii) To evaluate the high-energy behavior of scattering amplitudes, we make the Regge assumption that the amplitude is dominated by the contributions from an exchange in the crossed channel of these master trajectories. The residue functions automatically satisfy $U_W(6)$ invariance.

It appears that this Regge model will provide a reasonably restrictive theoretical framework for analysis

* On leave of absence from Imperial College, London, England.

[1] R. Migneron and K. Moriarty, Phys. Rev. Letters **18**, 978 (1967).

[2] Abdus Salam and J. Strathdee, Phys. Rev. Letters **19**, 339 (1967) (referred to hereafter as I); R. Delbourgo, Abdus Salam, M. A. Rashid, and J. Strathdee, Phys. Rev. **170**, 1477 (1968) (referred to hereafter as II).

[3] R. Delbourgo, M. A. Rashid, Abdus Salam, and J. Strathdee, in *Proceedings of the International Seminar in High-Energy Physics and Elementary Particles* (International Atomic Energy Agency, Vienna, 1965), p. 455.

of experimental data. Naturally, this theory will not provide any antidote to the obvious failures of conventional Regge techniques nor will it provide a fundamental answer to the unitarity difficulties which beset supermultiplet schemes. But it does give the possibility of building unitarity into the formalism as this is always done in Regge theory, i.e., mainly through the signature factor.[4] The new formalism will, however, certainly provide relations between presently used Regge residue parameters.

2. PARTIAL-WAVE EXPANSION IN $U(6)\otimes U(6)$

The basic ideas of the approach were described in I and II. Here we shall present a simplified version of the generalized expansion technique, proceeding by direct analogy with the conventional partial-wave expansion of the S matrix. The conventional partial-wave expansion can be understood either as a consequence of rotation invariance of the S matrix—and this of course is the deeper point of view—or, alternatively, as a *mathematical expansion* in terms of an appropriately chosen complete set of functions. It is this latter point of view that we wish to stress in this paper.

The rotation symmetry of the S matrix manifests itself in the following ways:

(a) Particles at rest group themselves into $(2J+1)$-component multiplets of $SU(2)_J$. (If the masses of the particles vary with J, one has a strong suggestion towards grouping them on a Regge trajectory.)

(b) A three-point function with all particles confined to the 0-3 plane shows helicity conservation:

$$\langle\lambda|T(E)|\lambda_1\lambda_2\rangle=\delta_{\lambda,\lambda_1-\lambda_2}T_{\lambda_1\lambda_2}(E). \tag{2.1}$$

(c) A four-point function with all particles also confined collinearly (forward scattering) shows net helicity conservation:

$$\langle\lambda_3\lambda_4|T(E)|\lambda_1\lambda_2\rangle=\delta_{\lambda_3-\lambda_4,\lambda_1-\lambda_2}T_{\lambda_3\lambda_4,\lambda_1\lambda_2}(E). \tag{2.2}$$

Suppose now that we are dealing with a nonforward scattering amplitude with the final particles rotated through an angle θ out of the 0-3 plane. We can always extract the angular dependence of $T(E,\theta)$ by expanding in a complete set of orthonormal (square integrable) functions as follows:

$$T(E,\theta)=\sum_n T_n(E)f_n(\theta).$$

The completeness of f_n means that a one-one correspondence between T_n and $T(\theta)$ exists. If we know nothing about the rotational invariance of the S matrix but simply that conditions (b) and (c) hold as empirical experimental facts, it is appropriate to choose the complete set of functions f_n to be the two-labelled function $d_{\lambda\lambda'}{}^J(\theta)$, satisfying $d_{\lambda\lambda'}{}^J(0)=\delta_{\lambda\lambda'}$; as one well knows, a class of such functions is given by the rotation functions of $SU(2)_J$. Thus one writes the mathematical expansion

$$\langle\lambda_3\lambda_4|T(E,\theta)|\lambda_1\lambda_2\rangle=\sum_J(2J+1)\times T_{\lambda_3\lambda_4,\lambda_1\lambda_2}{}^J(E)d_{\lambda_3-\lambda_4,\lambda_1-\lambda_2}{}^J(-\theta). \tag{2.3}$$

Expressing the summation as a Sommerfeld-Watson integral, one may tie in (c) with (a) and (b) in the well-known manner by proposing that $T_{(\lambda i)}{}^J(E)$ exhibits poles in the expansion-parameter J according to

$$\langle\lambda_3\lambda_4|T(E,\theta)|\lambda_1\lambda_2\rangle=\sum_{J,\lambda,\lambda'}\delta_{\lambda_3-\lambda_4,\lambda'}g_{\lambda_3\lambda_4}{}^J\times\frac{d_{\lambda'\lambda}{}^J(-\theta)}{E^2-m_J{}^2}g_{\lambda_1\lambda_2}{}^J\delta_{\lambda,\lambda_1-\lambda_2}. \tag{2.4}$$

Let us generalize. If the rotational symmetry $SU(2)_J$ is combined with $SU(3)$ to give a possible *rest symmetry* $U(6)\otimes U(6)$ and if $U(6)\times U(6)$ was known to be the symmetry of at least a part of the S matrix—a very strong assumption and certainly false for the exact S matrix—the symmetry would manifest itself in the following ways:

(a′) Physical particles group themselves in $U(6)\otimes U(6)$ multiplets.[5,6] (If the first few representations are known it would be natural to attempt to trace a Regge trajectory through them.)

(b′) Three-point functions exhibit W-spin conservation[7] (generalized helicity conservation—see Appendix). Thus

$$\langle W|T(E)|W_1W_2\rangle=\sum_\zeta\langle\zeta W|W_1W_2\rangle T_{\zeta W_1W_2}(E), \tag{2.1′}$$

where $\langle\zeta W|W_1W_2\rangle$ denotes the $U(6)_W$ Clebsch-Gordan coefficient which couples $D^{W_1}\otimes D^{W_2}$ to D^W. In general there is more than one independent coupling. It is therefore necessary to include a parameter ζ to distinguish among them.

(c′) Collinear scattering processes also exhibit $U(6)_W$ conservation:

$$\langle W_3W_4|T(E)|W_1W_2\rangle=\sum_{\zeta\zeta'W}\langle W_3W_4|\zeta'W\rangle\times T_{\zeta'\zeta}{}^W(E)\langle\zeta W|W_1W_2\rangle. \tag{2.2′}$$

[4] The absorption aspects of Regge theory arise, as is well known, mainly from the signature factor. This is because the Regge amplitude $(1\pm e^{i\pi\alpha})\beta(t)\hat{S}^{\alpha(t)}/\sin\pi\alpha$ is real (and violates unitarity) for real α and β, if the crucial signature factor is not included.

[5] R. Delbourgo, Abdus Salam, and J. Strathdee, Proc. Roy. Soc. (London) 284A, 146 (1965); B. Sakita and K. C. Wali, Phys. Rev. 139, B1355 (1965).

[6] R. Dashen and M. Gell-Mann, Phys. Letters 17, 142 (1965); H. Harari and H. Lipkin, Phys. Rev. 140, B1617 (1965); P. G. O. Freund, Phys. Rev. Letters 14, 803 (1965); R. Oehme, *ibid.* 14, 664 (1965).

[7] K. J. Barnes, Phys. Rev. Letters 14, 789 (1965); H. Lipkin and S. Meshkov, *ibid.* 14, 670 (1965); R. Arnold, Phys. Rev. 162, 1334 (1967).

(d′) Noncollinear four-point functions show conservation of coplanar symmetry $U(3)\otimes U(3)$ which has no analog for the smaller rest symmetry $SU(2)_J$.

If we accept only that (a′), (b′), (c′), and (d′) hold as empirical facts (at least to a fair approximation), we may adopt the mathematical expansion theorem attitude and express nonforward scattering amplitudes in terms of the *complete* set of suitably defined functions $d_{WW'}{}^N(\theta)$ as follows:

$$\langle W_3W_4|T(E,\theta)|W_1W_2\rangle = \sum_{N\zeta W\zeta'W'} \langle W_3W_4|\zeta'W'\rangle d_{W'W}{}^N(-\theta) \times T_{\zeta'W',\zeta W}{}^N(E)\langle\zeta W|W_1W_2\rangle. \quad (2.3')$$

To satisfy the boundary conditions (b′), (c′), and (d′), the suitable definition of $d_{WW'}{}^N(\theta)$ turns out to be that these functions are $U(6)\times U(6)$ rotation functions [$d_{WW'}{}^N(0)=\delta_{WW'}$; also $d_{WW'}{}^N(\theta)$ are diagonal in $U(3)\otimes U(3)$ labels subsumed in W] but this is indicental for our present purposes.

What exactly is the nature of the relevant Casimir invariant N of $U(6)\times U(6)$? The completeness notion used here requires that we sum over a *one-parameter* family of $U(6)\times U(6)$ representations $\mathfrak{D}^N$ since we are eliminating thereby a single angle θ.[8] Moreover, if the representations $\mathfrak{D}^N$ are nondegenerate in their $U(6)_W$ content, i.e., if a complete set of basis vectors can be labelled $|NW\rangle$, then the functions

$$d_{W'W}{}^N(-\theta)=\langle NW'|e^{i\theta J_2}|NW\rangle$$

are well defined. One may show that any square-integrable function defined over the interval $0\leqslant\theta\leqslant\pi$ and satisfying the appropriate boundary conditions at $\theta=0,\pi$ can be expanded in terms of the $d_{WW'}{}^N(\theta)$ if we characterize the representation D^N by, for example, the symmetrized $U(6)\times U(6)$ tensors $\phi_{\alpha_1\cdots\alpha_{N+B}}{}^{\beta_1\cdots\beta_{N-B}}$, where B denotes the baryon number and N takes the values $\frac{3}{2}B$, $\frac{3}{2}B+1$, $\frac{3}{2}B+2$, $\cdots$, i.e., N *is the quark number.*[9]

Returning to the expansion (2.3′) finally one ties in the property (a′) by assuming that $T^N(E)$ exhibits poles in the N Casimirs corresponding to $U(6)\otimes U(6)$ bound multiplets, thereby reducing expression (2.3′) to

$$\langle W_3W_4|T(E,\theta)|W_1W_2\rangle = \sum_{N\zeta W\zeta'W'} \langle W_3W_4|\zeta'W'\rangle g_{\zeta'W'W_3W_4}{}^N \times\frac{d_{W'W}{}^N(-\theta)}{E^2-m_N{}^2} g_{\zeta WW_1W_2}{}^N\langle\zeta W|W_1W_2\rangle. \quad (2.4')$$

This is the direct analog of (2.4). The rotation functions $d_{WW'}{}^N(\theta)=\langle NW|e^{-i\theta J_2}|NW'\rangle$ which make their appearance are generalized derivatives of the Gegenbauer $C_N{}^2$ (see next section) just like the $d_{\lambda\lambda'}{}^J(\theta)$, which are generalized derivatives of the Legendre. We can now pass to the Regge amplitude by making a Sommerfeld-Watson transformation:

$$\lim_{(\cos\theta\to-\infty)} \langle W_3W_4|T(E,\theta)|W_1W_2\rangle \sim \sum_{\zeta W\zeta'W'} \langle W_3W_4|\zeta'W'\rangle g_{\zeta'W'W_3W_4}{}^\alpha \frac{d_{W'W}{}^\alpha(-\theta)}{\sin\pi\alpha(E)} \times g_{\zeta WW_1W_2}{}^\alpha\langle\zeta W|W_1W_2\rangle, \quad (2.5')$$

where $\alpha(m_N{}^2)=N$ is the master trajectory function. This is of course the direct analog of the normal Reggeization procedure which yields

$$\lim_{(\cos\theta\to-\infty)} \langle\lambda_3\lambda_4|T(E,\theta)|\lambda_1\lambda_2\rangle \sim\sum_{\lambda\lambda'} \delta_{\lambda_3-\lambda_4,\lambda'}g_{\lambda_3\lambda_4}{}^\alpha \frac{d_{\lambda'\lambda}{}^\alpha(-\theta)}{\sin\pi\alpha(E)} g_{\lambda_1\lambda_2}{}^\alpha\delta_{\lambda,\lambda_1-\lambda_2}. \quad (2.5)$$

Note the very close correspondence between Eqs. (2.1)–(2.5) and (2.1′)–(2.5′). If we multiply expressions (2.5) or (2.5′) by the signature factors $(1\pm e^{i\pi\alpha})$, we shall be taking some account of unitarity in the sense that absorptive effects on the high-energy amplitude are incorporated through this.

3. ROTATION FUNCTIONS IN $U(6)\otimes U(6)$

Any further progress requires a practical knowledge of the $d_{WW'}{}^\alpha(\theta)$ functions which appear in (2.5′). This section is devoted to their computation and tabulation.[2,10] The first and most direct method would be to work directly in the basis $|NW\rangle$ and to determine the $d_{WW'}{}^N(\theta)$ by setting up differential equations for them. A second, less direct but more feasible, method which we shall use instead is to work in an auxiliary relativistic basis $|A_1\cdots A_N\rangle$ and to calculate the (M-function-like)

[8] This one-one count is easy to see for the orthogonal group. For example, in $O(4)$ the generalized helicity group is $O(3)$, while $O(2)$ plays the role of the coplanar group, which means that one encounters only the flipless amplitudes $T_{\lambda\lambda'}(\theta)$. The expansion technique then replaces λ and θ by the two Casimir labels j_0 and σ appropriate to $O(4)$. In detail,

$$T_{J\lambda J'}(\theta)=\textstyle\sum_{j0\sigma} T_{JJ'}{}^{j0\sigma}d_{J\lambda J'}{}^{j0\sigma}(\theta),$$

where $d(\theta)$ are the complete set of "rotation" functions for $O(4)$.

More generally, for the case of $O(\nu)$ the expansion theorem reads

$$T_{N_\nu-1N_{\nu-2}N_{\nu-1}'}(\theta)=\textstyle\sum_{N_\nu} T_{N_{\nu-1}N_{\nu-1}'}{}^{N_\nu}d_{N_{\nu-1}N_{\nu-2}N_{\nu-1}'}{}^{N_\nu}(\theta),$$

where N_ν stand for the Casimir operators of $O(\nu)$. Since $(N_\nu)=(N_{\nu-2})+1$, the one-one count is clearly exhibited. The same is true for our case of $U(\nu)\otimes U(\nu)$.

[9] It is of course not essential to employ the set of most degenerate representations characterized by the pair of quantum numbers N and B in defining the complete set of functions. However, this choice involves the least complication since the less degenerate representations of $U(6)\otimes U(6)$ are not completely labelled by the $U(6)_W$ quantum numbers and it would be necessary to formulate more involved criteria for picking out orthonormal sets of functions $d_{WW'}{}^N(\theta)$ from among all the matrix elements of $e^{-i\theta J_2}$. This course may be forced upon us, however, if physical particles cannot all be accommodated on the trajectories obtained by Reggeizing simply the quark number N.

[10] R. Delbourgo, J. Math. Phys. (to be published).

$d_{A_1\cdots A_N,B_1\cdots B_N}(\theta)$; passing to the standard basis via the transformation wave functions $\langle A_1\cdots A_N|NW\rangle$, we recover the canonical $d_{WW'}{}^N(\theta)$. There are several advantages in following this seemingly indirect path.

(i) Crossing complications that occur in the canonical basis are avoided. All one needs is to differentiate between particle and antiparticle wave functions $u_{A_1\cdots A_N}\times(NW)=\langle A_1\cdots A_N|NW\rangle$ and $v_{A_1\cdots A_N}(NW)=\langle A_1\cdots A_N|\bar{N}\bar{W}\rangle$, respectively, in passing from one channel to another.

(ii) Tied to (i) is the problem of kinematical constraints on canonical basis amplitudes $T_{(W_i)}$ in passing from one channel to another. In the M-function approach these constraints are automatic (after contraction over the wave functions) and need not be considered separately provided that the invariant amplitudes in M are kinematic singularity free.

(iii) The use of the relativistic basis $|A_1\cdots A_N\rangle$ permits us to discuss in a simple manner the case where the total four-momentum vanishes. Moreover, off-mass-shell continuations appear to be more straightforwardly carried out for M functions than for $T_{(W_i)}$.

(iv) *The most important advantage* of using the M-function approach is that symmetry-breaking prescriptions can be readily formulated, particularly the symmetry breaking which comes about through using physical masses of particles rather than mean masses of multiplets and which affects even the Clebsch-Gordan coefficients $(W_1W_2|W)$. This is not easy to do after one has passed to $T_{(W_i)}$.

The auxiliary basis appropriate to (a')–(d') is of course provided by the nonunitary representations[3] of $U(6,6)$. The interpretation of the supermultiplet condition (b') is that one is limited to couplings involving the $U(6,6)$ auxiliary fields $\psi_{A_1\cdots A_N}(p)$ and the momenta $q_A{}^B$ only, while (a') is assured by subjecting the $U(6,6)$ fields to subsidiary Bargmann-Wigner equations. (c') and (d') are natural consequences of applying these rules to open diagrams.

The $d(\theta)$ functions may be calculated by inserting a general pole contribution specified by the quark number N into the scattering diagram. Before carrying out the contraction over external wave functions one meets $d^N(\theta)$ with a certain number of $U(6,6)$ indices (*the number depending on the external particles alone*). It is these which we list below for some simple cases rather than the contracted forms $d_{WW'}{}^N$.

Take the case of meson ($B=0$) exchange first and various simple examples.

$A.\ (1,1)_{\frac{1}{2}p+q}+(1,1)_{\frac{1}{2}p-q}\to(1,1)_{\frac{1}{2}p+q'}+(1,1)_{\frac{1}{2}p-q'}$

$$d^N(\theta)=C_N{}^3(\cos\theta),\quad \cos\theta=-\hat{q}\cdot\hat{q}'+\hat{q}\cdot p\hat{q}\cdot p'/m^2. \tag{3.1}$$

$B.\ (1,1+(6,\bar{6})_A{}^B\to(1,1)+(6,\bar{6})$

There are two separate contributions to the amplitude corresponding to the canonical functions $d_{1\,1}{}^N$ and $d_{1\,35}{}^N$. The amplitude is therefore described by the general linear combination

$$[g_1g_A{}^B+g_2\partial/\partial q_B{}^A]C_N, \tag{3.2}$$

where

$$\nu(\partial C_N/\partial q_B{}^A)=(\Gamma_{-q'})_A{}^BC_{N'}-(\Gamma_{-q})_A{}^BC_{N-1'}, \tag{3.3}$$

$$\Gamma_{\pm k}=(1/4m^2)(\not{p}\pm m)\gamma\cdot\hat{k}(\not{p}\mp m). \tag{3.4}$$

$C.\ (1,1)+(6,1)_A\to(1,1)+(6,1)_B$

The linear combination here is modified to

$$[g_1\delta_A{}^B+g_2\partial/\partial q_B{}^A]C_N. \tag{3.5}$$

$D.\ (1,1)+(6,\bar{6})_A{}^B\to(1,1)+(6,\bar{6})_{A'}{}^{B'}$

This is a generalization of process B, the amplitude now containing a double derivative

$$[g_1q_A{}^B+g_2\partial/\partial q_B{}^A][g_1q'_{B'}{}^{A'}+g_2\partial/\partial q'_{A'}{}^{B'}]C_N. \tag{3.6}$$

The single differentiation formula has been written above; the double differentiation gives

$$\begin{aligned}\nu(\nu+2)(\partial^2C_N/\partial q'_{A'}{}^{B'}\partial q_B{}^A)&=(\Gamma_+)_{B'}{}^B(\Gamma_-)_A{}^{A'}C_{N+1}{}''\\&+[(\Gamma_{+q})_{B'}{}^{A'}(\Gamma_{-q'})_A{}^B-(\Gamma_{+q'q})_{B'}{}^B(\Gamma_-)_A{}^{A'}\\&-(\Gamma_+)_{B'}{}^B(\Gamma_{-qq'})_A{}^{A'}]C_N{}''+[(\Gamma_{+qq'})_{B'}{}^B(\Gamma_{-qq'})_A{}^{A'}\\&-(\Gamma_{+q})_{B'}{}^{A'}(\Gamma_{-q})_A{}^B-(\Gamma_{+q'})_{B'}{}^{A'}(\Gamma_{-q'})_A{}^B]C_{N-1}{}''\\&+(\Gamma_{+q'})_{B'}{}^{A'}(\Gamma_{-q})_A{}^BC_{N-2}{}'',\end{aligned} \tag{3.7}$$

where

$$\Gamma_\pm=(1/2m)(\not{p}\pm m), \tag{3.8}$$

$$\Gamma_{\pm kk'}=(1/8m^3)(\not{p}\pm m)\gamma\cdot\hat{k}(\not{p}\mp m)\gamma\cdot\hat{k}'(\not{p}\pm m).$$

Contraction of $\partial^2C/\partial q'\partial q$ over the external wave functions provides $d_{35\,35}{}^N(\theta)$. The calculations for more complicated $d(\theta)$ involving further derivatives have not as yet been carried out.

We now turn to the simple cases involving baryonic exchanges. For the single-quark family exchange $B=\frac{1}{3}$ there is the basic process

$C'.\ (1,1)+(6,1)\to(1,1)+(6,1)$

In this case we arrive at

$$(d^N(\theta))_A{}^B=(\Gamma_+)_A{}^BC_{N+1/2}{}'-(\Gamma_{+q'q})_A{}^BC_{N-1/2}{}',\qquad N=\tfrac{1}{2},\tfrac{3}{2},\cdots \tag{3.9}$$

relating to $d_{6,6}{}^N(\theta)$. On the other hand, for the more practical case of $B=1$ exchanges we must consider the basic process

$E.\ (1,1)+56,1)(1,1)+(56,1)$

Suppressing the six obvious multispinor indices

$$\begin{aligned}d^N(\theta)&=\Gamma_+\Gamma_+\Gamma_+C_{N+3}{}'''-3\Gamma_{+q'q}\Gamma_+\Gamma_+C_{N+2}{}'''\\&+3\Gamma_{+q'q}\Gamma_{+q'q}\Gamma_+C_{N+1}{}'''-\Gamma_{+q'q}\Gamma_{+q'q}\Gamma_{+q'q}C_N{}'''.\end{aligned} \tag{3.10}$$

The $d_{\bar{5}6,56}{}^N$ functions[10] which may be deduced from this have been given in detail elsewhere.

All these functions need to be multiplied by the threshold factor $(|q||q'|)^N$ which appears naturally in M-function calculations. This was shown explicitly in I. The Regge formulas therefore appear in the form (omitting signature factors) $\approx g^\alpha(t)g^\alpha(t)(|q||q'|)^\alpha d^\alpha/\sin\pi\alpha$, where the β's are reduced residues.

4. FEYNMAN TRAJECTORIES AND THEIR DECOMPOSITION INTO REGGE TRAJECTORIES

The master N trajectories [we shall sometimes refer to them as leaning Feynman towers since the particles on them correspond to the most degenerate tower

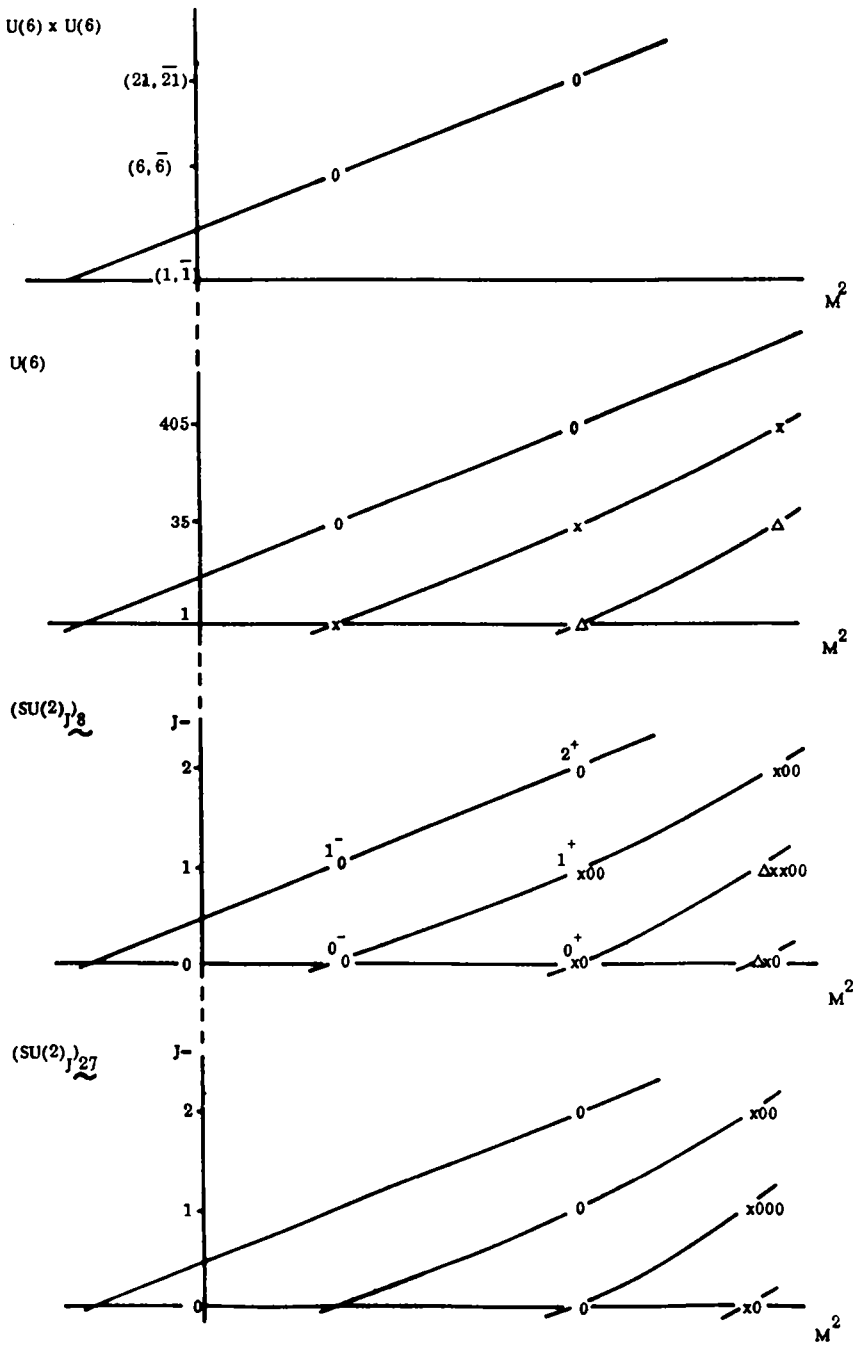


FIG. 1. The master boson trajectory decomposed into $SU(6)$ satellites (identified by 0, x, Δ, ..., etc.) which are further decomposed into $SU(3)$ pieces. Notice that there is more than one satellite trajectory of a given $SU(3)$ type. The symmetry breaking is expected to shift the trajectories from the positions shown. The known octets of $J^P=0^-$, 1^-, and 2^+ (and possibly 1^+) are shown in the third pattern.

studied by Feynman[11] for the noncompact $U(6,6)$ for mesons and baryons] contain all the relations between the J-Regge parameters. To see exactly what these relations turn out to be, we have to carry out the reduction chain $SU(6)\otimes SU(6)\to SU(6)\to SU(3)_F\otimes SU(2)_J$ of the master trajectories into the $SU(2)_J$ satellites for specific $SU(3)$ representations. Mathematically this reduction corresponds to the decomposition of particular $SU(3)$ components of the $SU(6)\otimes SU(6)$ rotation functions $C_N{}^3$ into the $SU(2)_J$ rotation functions $P_J=C_J{}^{1/2}$. The relevant formula is obtained from

$$C_N{}^\lambda(\cos\theta)=\sum_{\kappa\geq 0} a_{N\kappa}C_{N-2\kappa}{}^{\lambda'}(\cos\theta),$$

where the summation terminates at the background[12] and $a_{N\kappa}$ is a ${}_4F_3$ function (a sum of Γ-function ratios). For the simple case of the reduction

$$C_N{}^3=\textstyle\sum_\kappa a_{N\kappa}P_{N-2\kappa} \qquad (4.1)$$

the explicit formula was given in Eq. (15) of II. In the next section when we consider symmetry breaking, we shall need this reduction. In the M-function approach of Sec. 3, where all d^N's are expressed in terms of $C_N{}^3$ and its derivatives, it is just the formula (4.1) which is repeatedly needed.

To illustrate the consequences of this type of reduction graphically, let us plot a few satellite trajectories for the meson case. The master trajectory is shown in Fig. 1. It gives rise to the satellites shown in the lower diagrams. The rotation function $d_{WW'}{}^N$ pertaining to the master trajectories is a sum of rotation functions for all satellites. The general properties of these satellites have been noted in II. Here let us reemphasize the main physical points.

A. Parallel Satellites

If the symmetry were exact, all satellites would be parallel to the parent. Since empirically different $SU(2)$ *and indeed* $SU(3)$ trajectories are found to be roughly parallel (with the exception of the Pomeranchukon which may be a fixed pole) higher symmetry may provide the simplest explanation of this fact.

B. Residue Relations

In the asymptotic limit $P_J(\cos\theta)\approx(\cos\theta)^J$; it is clear from this that the leading satellite trajectory contained in the expansion (4.1) will dominate the scattering amplitude, e.g., the $SU(3)$ octet piece will show a dominance of the ρ trajectory over the π-trajectory contribution. On the other hand, the octet part of function $d_{35\ 35}{}^N$ automatically includes the contribution of both trajectories and the decomposition $C_N=a_{N0}P_N+a_{N1}P_{N-2}+\cdots$ shows how, for example, the ρ- and π-trajectory contributions emerge. This reduction provides group-theoretic relations between the Regge residues a_{N1}/a_{N0} automatically.

C. Symmetry Breaking

Since the high-energy behavior in Regge theory depends so critically on $\alpha(t)$, and in particular on the intercept $\alpha(0)$, it is evident that any mass shifts[13] produced by the symmetry breaking will shift the resultant satellites, and their asymptotic contributions will differ markedly from the exact symmetry predictions. This is in contrast to the effect of symmetry breaking for vertices where, barring certain exceptional cases, one hopes that symmetry breaking may be wholly accounted for just by change of kinematical factors, e.g., by using physical masses in the invariant couplings (and Bargmann-Wigner equations) rather than mean supermultiplet masses. To show how critical a role this trajectory shifting can play, take the example of pure **27** of $SU(3)$ exchange that occurs in a process like $K^-p\to\pi^++Y^-$, which shows no forward peak, and a high-energy behavior $E^{-3.5\pm0.7}$ corresponding to $\alpha_{27}(0)\approx-0.7$. Assigning the **27** as well as the 2^+ octet (f,A_2, K^{**},$\cdots$) to the same **405** of $SU(6)$, it is clear that an $SU(3)$-dependent mass shift between the **27** and the **8** of the order of no more than 300 MeV (without change of slope) can shift[14,15] $\alpha_8(0)$ from its value of about 0.4 down to $\alpha_{27}(0)\approx-0.7$.

[11] A. O. Barut, P. Budini, and C. Fonsdal, Phys. Rev. Letters **14**, 968 (1965); Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Letters **17**, 148 (1965) (the Feynman towers with the same content as the master trajectories were first presented in this paper); C. Fronsdal, in *Proceedings of the International Seminar in High-Energy Physics and Elementary Particles* (International Atomic Energy Agency, Vienna, 1965), p. 665; R. Delbourgo, Abdus Salam, and J. Strathdee, Proc. Roy. Soc. (London) **289A**, 177 (1966); Abdus Salam and J. Strathdee, Phys. Rev. **148B**, 1352 (1966).

[12] The exact form of the general formulas, which take into account also the background terms when N is complex, will be the subject of a further publication. The complicating point about such formulas is that the backgrounds of C^λ and $C^{\lambda'}$ occur at different places, namely, at $N=-\lambda$ and $N=-\lambda'$, respectively.

[13] F. Gürsey, A. Pais, and L. Radicati, Phys. Rev. Letters **13**, 299 (1964); M. Bég and V. Sigh, *ibid.* **13**, 418 (1964); H. Harari and M. A. Rashid, Phys. Rev. **143**, 1354 (1966).

[14] If $\alpha_{27}(0)$ is so critically shifted it is obvious that within the present Reggeization scheme, J. D. Jackson's analysis [Phys. Rev. Letters **15**, 990 (1965)] of Johnson-Treiman-like relations [J. D. Carter, J. J. Coyne, D. Horn, M. Kugler, H. Lipkin, and S. Meshkov, *ibid.* **15**, 373 (1965)] and his negative conclusion about $SU(6)_W$ predictions in the forward direction for **405** exchanges no longer applies. On the positive side, R. Arnold in Ref. 7 has considered processes not involving **27** exchange where the ρ_1A_2 trajectory dominates in a $U(6)\times U(6)\times O(3)$ scheme. He finds reasonable disagreement with experimental results if he assumes an $SU_W(6)$ symmetry between residues. Also, as Jackson himself noted, no account was taken in his analysis of mass differences due to symmetry breaking. As pointed out in (iv) of Sec. 3, the use of M-function formalism is superior for this reason to the direct W-spin formalism since it allows mass differences to be taken account of in the residues.

[15] Barut has given a one-parameter mass formula for mesons in the traceless $SU(3)$ form, $m^2=m_0{}^2+\lambda^2\{j(j+1)-\frac{1}{3}[I(I+1)-\frac{1}{4}Y^2-C]\}$, with $m_0{}^2=16\times10^4$ (MeV)^2 and $\lambda^2=28\times10^4$ (MeV)^2. The constant C which guarantees tracelessness depends on $SU(3)$ Casimirs and equals $+1$ for octets and 8/3 for 27-folds. The formula fits the known octets with accuracy and would predict a mass more than 300 MeV higher for the spin two doubly charged 27's. It is also worth remarking that the leading trajectory n ver contains **10**, $\bar{\mathbf{10}}$, **35**, $\bar{\mathbf{35}}$, or other non-self-conjugate multiplets of $SU(3)$. This means that the exchanges of such multiplets are definitely suppressed at high energies.

D. Mass Formulas and Trajectory Shifts

To take account of trajectory shifts on account of symmetry breaking, we need mass formulas, which in general may have the form[13]

$$M^2=M^2(N,J,F) \\ =M_0^2(N)+M_1^2(F)+M_2^2(F,J), \tag{4.2}$$

where F denotes the $SU(3)$ labels (including I and Y). To incorporate the trajectory shifts, go back to the expression of Sec. 2:

$$T=\int \frac{dN}{\sin\pi N}\frac{b^N C_N(\cos\theta)}{t-M^2(N)}.$$

One may replace C_N exactly by[16] $\sum a_{J\kappa}P_J$; if we further decide to incorporate symmetry breaking by replacing $M^2(N)$ by $M^2(N,J,F)$, we obtain

$$T\sim\sum_\kappa \int \frac{dJ}{\sin\pi(J+\kappa)}\frac{b^{J+\kappa}a_{J\kappa}P_J(\cos\theta)}{t-M^2(\kappa,J,F)}. \tag{4.3}$$

The trajectory function $J=\alpha(t,\kappa,F)$ [obtained from solving for J the equation $t-M^2(\kappa,J,F)=0$] now allows (1) for possible $SU(3)$ shifts given by $M_1^2(F)$ in (4.1) and (2) for departures from parallelism among the satellite trajectories arising from the $M_2^2(F,J)$ term. In keeping with our program, we shall not interfere with the residues $\beta_{J+\kappa}(t)$. Let us examine the simplified form of a mass formula (4.2), where

$$M_0^2(N)=NM_0^2,\quad M_2^2(F,J)=J(J+1)M_2^2(F); \tag{4.4}$$

i.e., the master trajectory in the N plane rises linearly. It is a simple matter to solve out for the trajectory function from $t-M^2=0$; we get

$$\alpha(t,\kappa,F)=[t-M_1^2(F)]\left(\frac{1}{M_0^2}+\frac{(2\kappa-1)M_2^2(F)}{M_0^4}\right) \\ -\kappa(\kappa-1)\frac{M_2^2(F)}{M_0^2} \tag{4.5}$$

to lowest order in M_2^2/M_0^2. To this order the trajectories remain linear but with modified slopes, exhibiting an $SU(3)$ mass shift which depends on which satellite we are considering.

Since at present we have no reliable theoretical means for computing mass formulas—except perhaps as tadpole effects or as estimates from second-order self-energy graphs written in the language of current algebra—we have to take the trajectory parameters from experiment. This is a weakness of the present scheme.

5. RELATIVISTIC ASPECTS OF $U(6,6)$

Just as for forward scattering of equal-mass particles the little group enlarges from $O(3)$ to $O(3,1)$, likewise here $U(6)\otimes U(6)$ enlarges to $U(6,6)$ itself. The $O(3,1)$ partial-wave analysis at $P_\mu=0$ which was originally carried out by Toller[17] can similarly be done here for $U(6,6)$. Following the method of Freedman and Wang,[17] one first shows, for a certain unphysical range of s, that one may deal with the compact group structure $U(12)$ rather than $U(6,6)$ so far as partial-wave analysis and Reggeization are concerned, continuing back later to physical values of s. Denoting the $U(12)$ rotation functions by $d_{NN'}^{\mathfrak{N}}(\theta)$, where $\mathfrak{N}$ and N stand, respectively, for the set of $U(12)$ and $U(6)\otimes U(6)$ Casimirs, one can make the expansion at $P_\mu=0$,

$$\langle N_3W_3,N_4W_4|T(\theta)|N_1W_1,N_2W_2\rangle \\ =\sum_{NWN'}\langle N_3W_3,N_4W_4|N'W\rangle \\ \times\langle N'W|T(\theta)|NW\rangle\langle NW|N_1W_1,N_2W_2\rangle \\ =\sum_{NWN'\mathfrak{N}}\langle N_3W_3,N_4W_4|N'W\rangle T_{N'N}^{\mathfrak{N}} \\ \times d_{N'WN}^{\mathfrak{N}}(\theta)\langle NW|N_1W_1,N_2W_2\rangle. \tag{5.1}$$

In the case of $O(3,1)$ or $O(4)$ the appropriate rotation functions are known to be $C_N^1(\cos\theta)$ and their derivatives. *For the $U(12)$ or $U(6,6)$ degenerate series one can show that they are proportional to $C_N^{11/2}$ and their derivatives.*

The expansion (5.1) holds at $P_\mu=0$. It can, however, be extended to the case $t=P^2\neq 0$ for all W-spin-conserving amplitudes, since $d_{NWN'}^{\mathfrak{N}}(\theta)$ provide appropriate expansion functions for this case as well. This is analogous to the expansion of general flipless amplitudes for all momentum transfers[18] using $O(3,1)$ rotation. An extension to W-changing amplitudes is possible, analogous to the $O(3,1)$ expansion proposed recently by a number of authors[18] for spin-flip amplitudes. These expansions correctly incorporate threshold effects and at the same time have the merit of automatically building the Toller parent-daughter phenomenon into the formalism even for $t\neq 0$. Thus a trajectory in the $U(6,6)$

[16] It is worth pointing out that the factors $a_{J\kappa}$ are factorizable. This is a general consequence of completeness relations of the type

$$\langle NW'|e^{-i\theta J_2}|NW\rangle=\sum_{\substack{J\kappa,J'\kappa' \\ (J+\kappa=J'+\kappa'=N)}}\langle NW'|J'\kappa'\rangle \\ \times\langle J'\kappa'|e^{-i\theta J_2}|J\kappa\rangle\langle J\kappa|NW\rangle \\ =\sum_{\substack{J\kappa \\ (J+\kappa=N)}}\langle NW'|J\kappa\rangle P_J(\cos\theta)\langle J\kappa|NW\rangle,$$

where the vectors $|J\kappa\rangle$ denote a basis for the representation $\mathfrak{D}^N$ which diagonalizes the angular momentum $\mathbf{J}^2$.

[17] M. Toller, Nuovo Cimento, 53A, 671 (1968); D. Freedman and J. M. Wang, Phys. Rev. 153, 1596 (1967); A. Sciarrino and M. Toller, J. Math. Phys. 8, 1252 (1967).

[18] R. Delbourgo, Abdus Salam, and J. Strathdee, Phys. Letters 25B, 230 (1967); R. F. Sawyer, Phys. Rev. 167, 1372 (1968); G. Cosenza, A. Sciarrino, and M. Toller, University of Rome Report, 1968 (unpublished).

Casimir $\mathfrak{N}$ plane gives rise to a series of parent and daughter trajectories in the $U(6)\otimes U(6)$ N plane—all of these daughters unfortunately being parallel to the parent.[19]

To see the complexion of these daughters, take the Feynman meson trajectory in $U(6,6)$ which, for this degenerate series, passes through the $U(12)$ representations **1, 143, 5940,** $\cdots$. From the $U(6)\otimes U(6)$ reduction of these multiplets

$$1=(1,1),$$
$$143=(6,\bar{6})+(\bar{6},6)+(1,1)+(1,35)+(35,1),$$
$$5940=(21,\bar{2}\bar{1})+(\bar{2}\bar{1},21)+(6,\bar{6})+(\bar{6},6)+(6,\bar{1}\bar{2}\bar{0})+(\bar{1}\bar{2}\bar{0},6)+(\bar{6},120)+(120,\bar{6})+(1,1)+(1,35)+(35,1)+(35,35)+(405,1)+(1,405),$$

one is led to sets of $U(6)\otimes U(6)$ trajectories, among which is the master meson trajectory considered earlier. The important relativistic aspects which emerge are as follows.

A. Gribov Doubling

The reduction of $U(12)$ multiplets into $U(6)\times U(6)$ multiplets produces pairs of the variety $(A,B)\oplus(B,A)$.[20] For example $(6,\bar{6})$ is accompanied by $(\bar{6},6)$; likewise (35,1) by (1,35).

To characterize this doubling, one may say that the states are populated equally by composites of quarks (6,1) and pseudoquarks (1,6) [also by antiquarks $(1,\bar{6})$ and antipseudoquarks $(\bar{6},1)$]. Even apart from Tollerization, this particular doubling should have been expected from the Gribov-Pomeranchuk-Okun phenomenon which even in conventional Reggeization schemes would lead one to expect that a Reggeized quark state should be accompanied by a pseudoquark state from MacDowell symmetry.[21] If composites of quarks exist, one should expect composites of pseudoquarks also to exist.

The important point to note about the Gribov doubling is that whereas for fermions it always leads to *parity doubling* [$(56,1)\rightarrow(1,56)$] (the two states have opposite parity) this is not necessarily the case for mesons [consider $(6,\bar{6})\rightarrow(\bar{6},6)$; parity of the two states is the same].

B. Parity Doubling for Mesons

In addition to Gribov doubling (which, as remarked above, does not lead to parity doubling for mesons), another peculiarly Toller-like phenomenon of parity doubling for mesons does take place. This is the doubling implied, for example, for the **143** by $(6,\bar{6})+(\bar{6},6)\Rightarrow(1,35)+(35,1)$. This is analogous (but not the same) as the parity-doubling phenomenon for Toller's theory of $SL(2,C)$ when for mesons one may expect parity degeneracy whenever the Lorentz quantum number M in Toller's notation does not equal zero. Perhaps one way to understand this new doubling is to remark that the chiral subgroup $U(6)\times U(6)|_{\gamma_5}$ is as equally a subgroup of $U(12)$ as nonchiral $U(6)\times U(6)|_{\gamma_0}$. As we have seen above, *for Reggeization* $U(6,6)$ and $U(12)$ possess completely interchangeable roles; one may start with either group and pass to the other by continuations in s and t variables. One may expect the theory therefore to exhibit doubling associated both with the chiral as well as nonchiral subgroups.

All this is not too clear at present. What we seem to have is that whithin an S-matrix approach, at the point $P_\mu=0$, one can resolve the old dilemma of chiral $U(6)\times U(6)$ being a symmetry at the same time as well as $U(6)\times U(6)$ nonchiral.

6. OUTLOOK

It must be admitted that it needs trepidation and courage to propose a theory of the type suggested here where the expectation is that higher symmetries may exhibit themselves best in giving a coherent description of Regge residues.[22] This is because, on superficial evidence, the major necessary condition for the theory—the existence of a string of higher supermultiplets lying on the master trajectories—seems unfulfilled. Unfortunately the situation in this regard may remain unchanged for a number of years.

The higher supermultiplets of $U(6)\times U(6)$ contain vast numbers of particles. The present rate of resonance identification, notwithstanding the heroic efforts of experimental colleagues, is slow. The situation is complicated further because, as has been shown by Horn, Lipkin, and Meshkov and Abramsky and King,[23] firstly, the higher $SU(3)$ multiplets contained in these supermultiplets are hard to produce in normal meson-baryon and baryon-baryon collisions and, secondly, most of these resonances do not possess two-body decays. Also, there is a great amount of mixing going on when resonances have the same quantum numbers. Indeed much theoretical work needs to be done to identify experi-

[19] This parallelism is of course one shortcoming of the formalism. It is important to distinguish the $U(6)\otimes U(6)$ satellites and the $U(6,6)$ daughters. The first are a consequence of the supermultiplet symmetry, the second a consequence of its relativistic enlargement.

[20] V. Gribov, L. Okun, and I. Pomeranchuk, Zh. Eksperim. i Teor. Fiz. 45, 1114 (1963) [English transl.: Soviet Phys.—JETP 18, 769 (1964)].

[21] S. MacDowell, Phys. Rev. 116, 774 (1959).

[22] This is the point of view which has consistently been emphasized from the first when the symmetries were suggested rather that their use without qualification. Thus, for the S matrix, the outlook was stated in Ref. 5 as follows: "with the effective baryon-meson and meson-meson vertices available, it is a trivial step to write pole approximations for the strong-interaction four-point processes. With this approximation as the starting point, all S-matrix techniques (like Mandlestam representation, Reggeization, analytic continuation both in angular momentum and unitarity spin) are available for determining the complete $\bar{U}(12)$ S-matrix theory. This is so because, as a rule, all that the S-matrix theory requires are Born approximations as the "input."

[23] D. Horn, H. Lipkin, and S. Meshkov, Phys. Rev. Letters 17, 1200 (1966); J. Abramsky and R. King, 20 1408 (1968).

mental situations where there is most chance of observing these higher multiplets.

In practical applications of the theory, one difficulty has been noted in Sec. 4. This is the difficulty associated with symmetry-breaking effects in mass formulas and the trajectory shifts these can produce, so that the trajectory parameters must at present be taken from experiment. A second difficulty is connected with the general Reggeization program. The Regge-pole model, even with its large number of parameters, has spectacular failures as well as successes. The failures have been attributed to kinematic effects, imperfectly understood so far, and to the fact that pion-exchange effects (perhaps on account of their exceptionally long range) appear less amenable to a Regge treatment and more to absorption or coherent-droplet models. The Reggeization scheme presented in the present paper will inherit the conventional kinematical structure. To be sure, though, there will be new features, like the threshold factor $(|q'||q)^N$ rather than the conventional factor $(|q'||q|)^J$, mentioned in Sec. 4, and the new zeros contained in $a_{N\epsilon}$ of Eq. (4.1) as well as the new features which will arise from a consideration of sense and nonsense phenomena anew in the present case.[24]

It is possible that a Toller-like program may provide here, as in conventional Regge theory, one way to define singularity-free amplitude. It is perhaps worth remarking that something mathematically similar to a Toller expansion of conventional amplitudes in terms of $O(4)$ rotation functions[18] is automatically included in our formalism, through the $U(2)\times U(2)$ subgroup of $U(6)$ $U(6)$. Even though $U(2)\times U(2)$ has a completely different physical significance from $O(3,1)$, the rotation functions for the two cases are identical. Whether this feature is enough to take care of all kinematic singularities automatically, we do not know. Only experience with the formalism can tell.

To expand on this point, it has been stressed before[3] that the $U(6,6)$ theory has two relatively disconnected features: First, the obvious one, it includes the internal symmetry $SU(3)$; second, and unfortunately the less emphasized but in our view the more important feature, it includes the extension of the space-time Lorentz structure $SL(2,C)$ to the bigger (perhaps conformal) structure $U(2,2)$. This extension $U(2,2)\approx O(4,2)$ increases the number of "space-time" Casimirs from the two well-known ones of $SL(2,C)$ to three of $U(2,2)$. It was pointed out in Ref. 3 that the empirically well-established proportionality of electric and magnetic form factors of the proton is a direct consequence of this particular extension of $SL(2,C)$ space-time group to the $U(2,2)$ group. Thus, even if $SU(3)$ was a badly broken symmetry or if it was conclusively established that all hadron resonances make up only the 8's and the 10's of $SU(3)$ and never any other multiplet, it would still, in our view, make better dynamical sense for Reggeization ideas to make a partial-wave analysis using the $U(2,2)$ extension of space time structure[25] [in practice in terms of functions $C_N{}^1$ and their derivatives corresponding to the little group $U(2)\times U(2)$]. Thus the first logical step in Reggeization of higher symmetries is to consider Reggeization of $U(2)\times U(2)$; this will give baryon and meson trajectories with content similar to those derived from noncompact groups by Barut and Kleinert[26]; next, one may include isospin and extend the symmetry to $U(4)\times U(4)$ and, finally, with the inclusion of $SU(3)$ to $U(6)\times U(6)$. The kinematic factors $a_{NJ'}$ arising from the decomposition of the relevant $C_{N'}$ to $C_J{}^\lambda = a_{NJ'}P_J$ for each assumed symmetry would be different. [For $U(\nu)\otimes U(\nu)$ symmetry the rotation functions are $C_N{}^{\lambda\nu}\times(\cos\theta)$; for $U(2\nu)$ they are $C_N{}^{\nu-1}(\cos\theta)$; and for $O(\nu)$ they are $C_N{}^{\lambda\nu-1}(\cos\theta)$.]

Hopefully, experiment may distinguish between the various possibilities which correspond to the successive chains of symmetry breaking. One of the important parameters relevant to this distinction is the F/D ratio[27]; it is a mathematically fascinating problem to compute the F/D ratio along the $SU(3)$ 8-projection of the Feynman trajectory. Other problems are a better understanding of the mathematical expansion theorem for the case of less degenerate series, a simpler procedure for computing the relevant d^N functions, and, most critical of all, a reliable mass formula for use in (4.3).

APPENDIX

There are at least four formulations of $U(6,6)$ and its subgroup symmetries known to the authors.[2,28] Though their relative merits are hotly debated,[29] all of them unfortunately suffer from one shortcoming or another. All approaches do at least agree on the subgroup hierarchy of Sec. 2 as representing the maximal possible invariance attainable. To describe the approaches and their interrelations, let us briefly recall the group structure they use in order to explain the detailed differences. To begin with, there is the $U(6,6)$ algebra which is isomorphic to the algebra generated by the 16 Dirac matrices γ multiplied into nine $SU(3)$ matrices T^i:

$$(1,\sigma,\gamma_5,\gamma_5\sigma,\gamma_0,\gamma_0\sigma,\gamma_0\gamma_5,\gamma_0\gamma_5\sigma)\times T^i.$$

[24] These latter (not studied so far) present fascinating problems; the mysterious vanishing of a number of residues in conventional theory may possibly find a *kinematical* explanation in the present formalism. This may not be surprising if one remembers that the extended kinematics of this formalism is an expression of the dynamics of hadron physics.

[25] We summarize here the results on rotation functions; for the most degenerate representations the rotation functions correspond to derivatives of $C_N{}^{\nu-1/2}$ for $U(2\nu)$ and $C_N{}^{\lambda\nu}$ for $U(\nu)\times U(\nu)$ groups. We conjecture that the same Gegenbauer polynomials and their derivatives occur for all other representations of the relevant groups.

[26] A. O. Barut and H. Kleinert, Phys. Rev. **160**, 1149 (1967).

[27] This has recently been emphasized by P. N. Dobson, Jr., Phys. Rev. **163**, 1619 (1967).

[28] S. Coleman, Phys. Rev. 138, B1262 (1965), has listed a still larger number of variants.

[29] See, for example, Y. Ne'eman, *Algebraic Theory of Particle Physics* (W. A. Benjamin, Inc., New York, 1967) and P. T. Matthews, Nature **217**, 197 (1968), review of this book.

(The Lorentz subalgebra is generated by σ and $\gamma_5\sigma$.) Four translations P_μ are adjoined to $U(6,6)$, whose commutation property is obtained through the isomorphism $P_\mu \doteq \gamma_\mu$. For processes involving one (timelike) vector $P_0 \doteq \gamma_0$, the subgroup of $U(6,6)$ which commutes with γ_0 is the "little" group $U(6)\otimes U(6)$ which consist of $(1,\gamma_0,\sigma,\gamma_0\sigma)T^i$. Collinear processes confined to the 0-3 plane require the "lesser" group which commutes with the pair of vectors γ_0 and γ_3; this is $U(6)_W$ and consists of $(1,\sigma_3,\gamma_0\sigma_1,\gamma_0\sigma_2)T^i$ [for the Lorentz case the analogous subgroups are $SU(2)_J$, consisting of σ and the helicity group $U(1)$ consisting of σ_3 alone]. *W-spin is thus the generalized helicity of $U(6,6)$*. Finally, there are the coplanar processes confined to the 013 subspace whose "least" group is $U(3)\otimes U(3)$ made up of $(1,\gamma_0\sigma_2)T^i$; this has no analog in the Lorentz group case.

So much is common ground. However, the four approaches differ in the concrete realizations which they give to the generators of $U(6,6)$ and the way the translations P_μ are handled.

(1) First, there is the simple field-theoretic approach[30] based on a Lagrangian formulation of $U(6)\otimes U(6)$ multiplets, e.g., the quark Lagrangian $\mathfrak{L}=\bar{\psi}(i\gamma\cdot\partial-m)\times\psi+g(\bar{\psi}\psi)^2$ or a more complicated Lagrangian constructed from the $U(6,6)$ multispinors. In this formulation the mass and interaction terms are $U(6,6)$-invariant whereas the kinetic energy terms of the type $\bar{\psi}i\partial\psi$ are not.[6] Evidently open diagrams and their sums do possess the hierarchy of little group symmetries (even if derivative interactions are included), whereas closed loops are not likely to preserve these. If a Regge pole is pictured as an infinite sum of pole diagrams[31] the hierarchy of symmetries survives. However, inclusion of two-particle or more intermediate states, i.e., imposition of unitarity, breaks the chain through the symmetry breaking introduced by closed loops.

(2) The second approach was suggested by a number of authors[11] and developed in particular by Fronsdal and his collaborators. Here the full noncompact $U(6,6)$ may be taken as a rest symmetry with the consequence that there must exist an infinity of particle states corresponding to representations of $U(6)\otimes U(6)$ all *having the same mass*. The subgroup hierarchy provides *exact* invariance groups for the relevant processes; unitarity also is exactly satisfied but only in the mass degenerate limit—as soon as mass differences are introduced between different particle states unitarity disappears. It is clear that Reggeization of approach (1) and its interpretation as a summation over an infinity of particle states brings closer together approaches (1) and (2).

(3) The third approach is based on current algebras[6] and is wide enough to encompass either (1) or (2). Unhappily, there exists no model, however idealized, for which the charges defined from the full set of $U(6)\otimes U(6)$ currents are conserved.

(4) The last approach is the inhomogeneous $U(6,6)$ theory of Bell and Rüegg and Charap, Matthews, and Streater,[32] which adjoins 143 momenta to $U(6,6)$. Before specializing to four physical momenta the subgroup hierarchy, as well as unitarity in a generalized partial-wave expansion, emerge as *exact* consequences of the theory; also, one may write equations of motion for the nonunitary finite-dimensional representations of $U(6,6)$ since one is dealing with a 143-dimensional Poincaré group. In the physical limit of four-momenta surviving from among the 143, the equations of motion of approach (1) are obtained. One could write if one wished Majorana type equations in the (144) space for infinite-dimensional representations of $U(6,6)$ to give a physical particle spectrum. The unresolved difficulty of this approach is the definition of a sensible (stereographic) limit whereby the 143-dimensional space maps onto physical four dimensions.

In Sec. 2 we have tried to formulate yet another viewpoint by accepting the subgroup hierarchy as empirical input. We have worked with just the conventional S-matrix setup in the physical space of four dimensions. We have made a partial-wave analysis based on the existence of a complete set of functions in terms of which $S(\theta)$ can be expanded—a purely mathematical procedure which must always succeed provided the weak statement of the input hierarchy of subgroups is guaranteed by the choice of the expansion functions. The full symmetry of the S matrix under the higher group is not needed.

[30] R. Delbourgo, M. A. Rashid, Abdus Salam, and J. Strathdee, Proc. Roy. Soc. (London) **285A**, 312 (1965).
[31] L. Van Hove, Phys. Letters **24B**, 183 (1967).
[32] J. Bell and H. Rüegg, Nuovo Cimento **39**, 1166 (1965); J. Charap, P. T. Matthews, and R. F. Streater, Proc. Roy. Soc. (London) **290A**, 24 (1966).

Weak and electromagnetic interactions

By Abdus Salam

Imperial College of Science and Technology, London, England, and International Centre for Theoretical Physics, Trieste, Italy

One of the recurrent dreams in elementary particle physics is that of a possible fundamental synthesis between electromagnetism and weak interaction. The idea has its origin in the following shared characteristics:

1. Both forces affect equally all forms of matter-leptons as well as hadrons.
2. Both are vector in character.
3. Both (individually) possess universal coupling strengths.

Since universality and vector character are features of a gauge theory those shared characteristics suggest that weak forces just like the electromagnetic forces arise from a gauge principle. There is of course also a profound difference: electromagnetic coupling strength is vastly different from the weak. Quantitatively one may state it thus: if weak forces are assumed to have been mediated by intermediate bosons (W), the boson mass would have to equal from 50 to as large as 137 M_p, in order that the (dimensionless) weak coupling constant $g_w^2/4\pi$ equals $e^2/4\pi$.

I shall approach this synthesis from the point of view of renormalization theory. I had hoped that I would be able to report on weak and electromagnetic interactions throughout physics, but the only piece of work that is complete is that referring to leptonic interactions, so I will present only that today. Ward and I worked with these ideas intermittently (1, 2), particularly the last section on renormalization of Yang-Mills theories. The material I shall present today, incorporating some ideas of Higgs & Kibble, was given in lectures (unpublished) at Imperial College. Subsequently I discovered that an almost identical development had been made by Weinberg (3) who apparently was also unaware of Ward's and my work.

The renormalization point of view has become increasingly more and more important as time has gone on. In fact, if I have heard the reports rightly, the major activity in the U.S.A. last year, at least according to the rumours we heard about it, was the attempt to renormalize the radiative corrections to beta-decay, particularly at Princeton, by Gell-Mann, Goldberger, Kroll, Low and others. I will not say that the problem of renormalization of weak interactions, together with their radiative corrections, has acquired the same

Reprinted from Elementary Particle Theory – Relativistic Groups and Analyticity, ed. N. Svartholm.
© John Wiley & Sons, 1968, pp. 367–377.

importance now as in 1947 it had for electrodynamics in connection with the Lamb shift, but I think it is a pretty serious problem, which must be treated at a fundamental level.

Today it is not the renormalization to second order that one is speaking about. I have in mind a renormalization program that should apply for the complete theory just as the program applied for electrodynamics itself. From this point of view we shall see that this unification of electromagnetism and weak interactions becomes more and more of a necessity. Further, it places a large number of restraints on the type of theory one can have.

For leptonic weak interactions the basic problem is of course that one is dealing with vector and axial vector interactions of spin half particles. The four-fermion interaction has not the slightest vestige of renormalizability. We thus consider intermediate boson theory and what we wish to do is to formulate the theory such that the photon appears as the neutral intermediate boson. Before I do this I shall briefly review what we know about the renormalization problem of vector mesons. As is well known, the lack of renormalizability of the theory arises from massive vector mesons from the $k_\mu k_\nu/m^2(k^2-m^2)$ term in the propagator.

(*a*) Consider *neutral vector* mesons, where there is hope that currents can be conserved. In this case Feynman's old proof, in which he showed that this $k_\mu k_\nu$ term latches on to the currents J_μ and J_ν at either of the vertices and disappears because of the conservation law, applies. The theory is renormalizable even if the vector meson is massive, as was first shown by Matthews. Thus current conservation must be preserved at all costs as a minimum requirement.

(*b*) Clearly, for an *axial* current the mass of the source Fermi particle (electron or meson), which makes the current non-conserved, will destroy any hope of renormalizability. Therefore the second restriction is that we must not have a fundamental Fermi mass in the theory.

(*c*) If one is dealing with charged vector or axial mesons the currents must include these particles themselves. To construct a conserved current this time we must have a recipe which includes contributions of the vector and axial vector mesons themselves within the current. The only known method for constructing such currents is the Yang-Mills recipe which arises when one is dealing with gauge symmetries associated with non-abelian symmetry groups.

(*d*) Now what about the renormalizability of Yang-Mills theories? Around 1962 the following was proved. Take $U(3)$ symmetry; then out of the, say, nine currents of $U(3)$, three of these, J_0, J_3 and J_8, will allow the corresponding $k_\mu k_\nu/m^2$ terms in the propagator to be transformed away. But, for the remaining currents the corresponding meson propagator must contain these terms and they cannot cancel *even though the currents are conserved.* Thus Yang-Mills

theory with *non-zero meson mass is not renormalizable even if the currents are fully conserved.*

Therefore the only hope for vector mesons is (1) a Yang-Mills theory but also (2) that mesons should be massless. *To state it more generally, it is not current conservation that is the criterion of renormalizability, but gauge invariance.* The mass of the vector meson is not gauge invariant even though it does not interfere with current conservation. Nothing that will destroy the gauge invariance is allowed. Thus the only hope for achieving a renormalizable theory involving vector currents is that the mass of the meson should come through a spontaneous symmetry breaking. Further, since the second source of lack of (axial) gauge invariance is going to be the Fermi mass term, what would be ideal is if the Fermi mass term could also come from spontaneous symmetry breaking and preferably the *same* symmetry breaking. The Fermi mass coming from symmetry breaking was an idea started by Nambu, and has been worked on particularly by Johnson, Baker and Willey in electrodynamics. One adds the mass term in the free part of the Hamiltonian, sets up an interaction representation, computes the self-mass and sets it equal to the physical mass. This gives a self-consistent equation. For the meson case the propagator has no $k_\mu k_\nu/m^2$ terms but just $\delta_{\mu\nu}/(k^2-m^2)$ and the theory from the outset is renormalizable. (See for more details ref. 2.)

A second proposal and a related one is not to do this rather brutal addition and subtraction of mass terms but to work more gently. This is the method of letting the vector mesons interact with a set of scalar particles and to let them acquire physical masses by assuming self-consistently that these scalar particles possess non-zero expectation values. Consider, for example, the charged mesons ϕ and ϕ^* interacting with a neutral vector meson. Compute self-consistently from the theory an expectation value $\langle\phi_0\rangle$ for the ϕ field. The term $e^2\phi^*\phi A_\mu^2$ in the Lagrangian now appears to have a piece $e^2(\phi_0)^2 A_\mu^2$ which acts like the vector meson mass.

Now all this is beautiful, but there was the suspected difficulty with this theory which held people back for a long time. This was the fear of Goldstone mesons. What one feared was that whenever you have such a theory Goldstone boson sits like a snake in the grass ready to strike. According to Goldstone's theorem a number of massless particles must arise in any such theory. To see these massless objects it is simplest to take the same example as before of charged scalar meson; set

$$\phi = \varrho e^{i\theta}, \quad \phi^* = \varrho e^{-i\theta} \tag{1}$$

The Lagrangian equals

$$\mathcal{L} = \partial_\mu\phi^*\partial_\mu\phi + m^2\phi^*\phi = (\partial_\mu\varrho)^2 + \varrho^2(\partial_\mu\theta)^2 + m^2\varrho^2 \tag{2}$$

It clearly has a massless field θ. It is this field θ which for spontaneous symmetry-breaking theories appears as the Goldstone boson.

A big advance in resolving this difficulty has recently been made but I do not think that many people are aware of it. This is due to Higgs and, following him, Kibble who have shown that, true, there may be massless scalar mesons when you have done the spontaneous symmetry breaking. *But if you have a gauge theory* such objects can be transformed away as gauges and they have no coupling to the physical particles. To see this let us go back to the same example. Consider the Lagrangian

$$\mathcal{L} = (\partial_\mu - ieA_\mu)\phi^*(\partial_\mu + ieA_\mu)\phi + F_{\mu\nu}F_{\mu\nu} + m^2\phi^*\phi \tag{3}$$

Now write

$$(\partial_\mu + ieA_\mu)\phi = e^{i\theta}(\partial_\mu \varrho + ieA'_\mu \varrho) \tag{4}$$

where $A'_\mu = A_\mu + \partial_\mu\theta$. Clearly the term $\partial_\mu\theta$ appears in such a way where it can be transformed away through the gauge invariance, and no trace of the θ field is left. Also, if we assume that the expectation value of ϱ could be non-zero, i.e. $\varrho = \langle\varrho_0\rangle + \varrho'$ where ϱ' is the quantized part of ϱ, then the term $e^2\varrho^2A_\mu^2$ now has a piece $e^2(\varrho_0)^2A_\mu^2$, i.e. the vector meson acquires a mass term in this particular gauge specified by $\langle\varrho_0\rangle$.

What has happened? We started with the two fields ϱ and θ. One of the fields, ϱ, I can use for spontaneous symmetry breaking. The other field was massless. The symmetry breaking gave mass to the vector meson and the massless objects got transformed away. So you have the best of all possible worlds in this theory. Since the mass term has come in gently through symmetry breaking, our claim is that the theory is renormalizable. It clearly is in the original formulation with ϕ, ϕ^* fields.

Let us now make use of this in the case of weak and electromagnetic interactions. I will try to get a Yang–Mills theory of weak and electromagnetic interactions combined because that is the only way of keeping currents conserved. Since there is a mass of the fermion, the Yang–Mills theory will want to be broken spontaneously. If the same symmetry breaking can be arranged to give masses to the charged intermediate bosons which mediate weak current but not to the neutral meson representing the electromagnetic field we would have removed the major objection to the unification of weak and electromagnetic fields of having in one multiplet massive and massless objects.

The application of the Yang-Mills ideas to weak and electromagnetic interaction has been carried through by Gatto, Ne'eman, Ward, Salam and recently Weinberg. I will write it in an $SU(3)$ notation and I hope that this will prove to be more than a notational nicety. What you do is that you have a four-

component neutrino, a negative electron and positive muon, and define the lepton triplet

$$L=\begin{pmatrix}\nu\\ e^-\\ \mu^+\end{pmatrix} \tag{5}$$

Consider I-spin and V-spin subgroups of $SU(3)$ where I define the scalars in I-space and V-space as

$$I_0=T^8,\ V_0=\frac{\sqrt{3}}{2}T^3-\tfrac{1}{2}T^8 \tag{6}$$

The lepton charge equals

$$Q_l=\begin{pmatrix}0 & & \\ & -1 & \\ & & 1\end{pmatrix} \tag{7}$$

Notice that the hadronic charge in the same notation is

$$Q_h=T^3+\frac{i}{\sqrt{3}}T^8=\begin{pmatrix}\frac{2}{3} & & \\ & -\frac{1}{3} & \\ & & -\frac{1}{3}\end{pmatrix} \tag{8}$$

so that Q_h and Q_l are "orthogonal" combinations in T^3, T^8 space. Define right and left particles, conventionally,

$$\psi_{L,R}=\tfrac{1}{2}(1\pm\gamma_5)\psi \tag{9}$$

and note that

$$\begin{aligned}I_{2L}-V_{2R}&=\tfrac{1}{2}(Q_l+3\gamma_5 Q_h)\\ I_{0L}-V_{0R}&=\tfrac{1}{2}(Q_l-\gamma_5 Q_h)\end{aligned} \tag{10}$$

Now the weak interaction can be written as

$$\mathcal{L}_{\text{weak}}=[(\bar{e}^-\nu)_L+(\bar{\nu}\mu^+)_R]\,W^-+\text{h.c.}=(I_{L1}+V_{R1})\,W_1+(I_{L2}-V_{R2})\,W_2 \tag{11}$$

where the intermediate bosons are W^+ and W^-. Note that the anti-neutrino right come in with muon and neutrino left with the electron so that all four components of the neutrino field are used up. This is a minor merit of the formulation. If one has a theory of symmetries it is undesirable to have to work with two-component neutrinos and four-component electrons and muons, because in some conceptual limit you would like to transform them into each other.

To complete the Yang-Mills gauge group we must introduce the third component of the vector W, adding on the term $(I_{3L} - V_{3R})\, W_3$ to the Lagrangian. Ideally one would have liked this to be nothing but the electromagnetic field. Unfortunately the $I_{3L} - V_{3R}$ term contains the $\gamma_5 Q_h$ piece which has nothing to do with electromagnetism. One must therefore introduce in this theory an additional neutral vector meson W_0 and add in a term

$$\mathcal{L}_{\text{neutral}} = (I_{3L} - V_{3R})\, W_3 - \sqrt{3} \tan\theta (I_{0L} - V_{0R})\, W_0 \tag{12}$$

to the Lagrangian (11). The factor $\sqrt{3} \tan\theta$ gives an arbitrary scale factor between the two neutral currents. In terms of the quantities

$$A^0 = W_0 \cos\theta + W_3 \sin\theta$$

$$X^0 = -W_0 \sin\theta + W_3 \cos\theta \tag{13}$$

one can write the neutral currents as

$$(\cos\theta Q_X X^0 + \sin\theta Q_1 A^0) \tag{14}$$

where

$$Q_X = \tfrac{1}{2} \sec^2\theta (3Q_h\gamma_5 + (2\cos\theta - 1)\, Q_1) \tag{15}$$

Identify A^0 with the Maxwell field. In addition one has to buy—as about the minimum price for this unification—another field X^0 which is neutral and has the current Q_X. Note that the current Q_X is diagonal in all the particles and does not transform electrons to muons.

Let us not worry about muon mass for the moment and concentrate on the electron. We shall now introduce one common source of breaking the symmetry introduced above which will give (*a*) mass to the electron (*b*) mass for W^+, W^- and for X^0 (but not to A^0). To bring out the structure of this term let us rewrite the symmetry introduced above in a more familiar language, define a lepton hypercharge and write

$$\tfrac{1}{2}Y = -\sqrt{3}\, T^8 \tag{16}$$

All that I described above in the $SU(3)$ notation can be re-expressed in the following way: we make a doublet out of the electron-left and neutrino-left with hypercharge -1 and let the electron-right correspond to $I=0$ with $Y=-2$. This can be indicated in this way

	I_3	Y	$Q = I_3 + \frac{1}{2}Y$
V_L^0	$\frac{1}{2}$	-1	0
e_L^-	$-\frac{1}{2}$	-1	-1
e_R^-	0	-2	-1

To bring out the analogy with hadron physics, it would be as if we were dealing with Ξ and Ω^-.

Write

$$\mathcal{L}_m = m\bar{e}_R e_L + \text{h.c.} \overset{\text{analogy}}{=} f\bar{\Omega}^-(\langle\mathcal{K}\rangle\tau_2\Xi) + \text{h.c.}; \quad \mathcal{K} = \begin{pmatrix} K^0 \\ K \end{pmatrix} \tag{17}$$

Since I would now want to get this mass term as an expectation value of some scalar entity, I have written it in a suggestive form in analogy with a sort of $\Delta I = \frac{1}{2}$ rule for I-spin breaking. To do this I have exhibited an analogous interaction of Ω^- and Ξ particles. The idea is that we introduce into the theory an object which transforms like the $\bar{K}$-meson in *leptonic I-space* and that this object—or rather the K_{01} component of it—possesses a non-zero expectation value, $f\langle K_{01}\rangle m$. The $\bar{K}$-meson thus introduced must itself have a gauge interaction plus a mass term (plus a four-field interaction to allow it to acquire a non-zero expectation value). Thus its Lagrangian is

$$\mathcal{L} = (\partial_\mu - ieW_\mu)\mathcal{K}^+(\partial_\mu + ieW_\mu)\mathcal{K} + m^2\mathcal{K}^+\mathcal{K} + \lambda(\mathcal{K}^+\mathcal{K})^2 \tag{18}$$

If one does a very simple-minded calculation the expectation value of K^0 comes out of the magnitude

$$\langle K_{01}\rangle \approx m^2/2\lambda \tag{19}$$

where m is the mass of the K-particle. We now use the Higgs ansatz and go into a special gauge where all objects except K_{01} are transformed away. Higgs would tell us that of the four real fields contained in $\mathcal{K}$, K_{01} to which we have hopefully assigned the spontaneously introduced non-zero expectation value, will survive in the Lagrangian while K_{02}, K^+ and K^- can be transformed away. What is further remarkable is that this will give mass terms proportional to $e^2\langle K_{01}\rangle^2$ to the charged intermediate bosons W^+, W^- and to X^0 but not A^0. The simplest way to see this is to remark that the vector meson field corresponding to the symmetry which survives (electric charge conservation) started with zero mass and will end up in the special gauge with zero mass. There is only one symmetry which we are preserving in the whole business and that is the symmetry corresponding to the electromagnetic field. We had no mass for this field to start with and no mass when we finish. The other three are the symmetries we break; the corresponding vector fields acquire masses which are $e^2\langle K_{01}\rangle^2$. The whole thing has therefore worked out in a unified and beautiful manner; a renormalizable theory of massive vector mesons, in which electromagnetism is built in as part of the theory and the Lagrangian is perfectly symmetrical. Only when one goes to a special gauge, where

$$\mathcal{K} = e^{i(\Pi\cdot\theta/\varrho)\gamma_5}\begin{pmatrix} \varrho \\ 0 \end{pmatrix} \tag{20}$$

and where $\langle\varrho\rangle = K_{01}$, only in this gauge does the Lagrangian look ugly.

Like all suggestions in physics this one too has some associated difficulties. The first difficulty of course is "what shall we do with this expectation value?" We would like the mass of the vector meson

$$m_V^2 = e^2\langle K_{01}\rangle^2 \tag{21}$$

to be larger than 50 BeV or so to get the weak interaction constant to equal the electromagnetism. If we want the coupling constant e to be the electromagnetic coupling constant, then $\langle K_{01}\rangle$ must be of the order of 500–1000 BeV. We can, of course, arrange for such a large value of $\langle K_{01}\rangle$ by making the mass m in front of the term $\mathcal{K}^+\mathcal{K}$ large. But this is really shifting the problem of a large mass for the intermediate boson to a (still) larger mass for the scalar K-meson.

This is not so pleasant. There may be another way out. Suppose we were using similar ideas for the baryon case, with something like a fundamental quark field which we assume to be very massive. If such a quark mass were also to arise from spontaneous symmetry breaking, the scalar object similar to what we called the K-meson above, coupling to such quarks, would need to possess an enormous value for its vacuum expectation value. Such a scalar object would couple with the intermediate bosons just like the K-particle above and give them mass. It may be more pleasant to find the source of large vector boson mass thus through its association with baryons rather than leptons. The point is that since weak bosons couple both to leptons *and* baryons, their large mass may come from the baryon side rather than leptons since all contributions of the type $\langle K_{01}\rangle^2$ which determine both the baryon masses and boson masses come additively in the latter.

All I am trying to say is that you must have a complete theory. And one must not be allowed to break symmetries in the fundamental Lagrangian ad hoc if vector meson renormalizability is to be preserved. I feel that this is a line worth pursuing just because it is so restrictive.

Now some remarks on the construction of a similar theory for hadrons. This Ward and myself tried very hard to do in 1964 but we failed. The main source of the failure was always the Cabibbo angle. One does not want the neutral current introduced into the theory to be strangeness changing. In every theory with a non-zero Cabibbo angle this will inevitably happen.

One line we tried was to use the same combination of I_L and V_R for the hadrons with two types of intermediate bosons

$$[I_L \cos\theta + V_R \sin\theta]\, W_1 + [-I_L \sin\theta + V_R \cos\theta]\, W_2 \tag{22}$$

as far as hadrons are concerned. Let only W_1-particles interact with the leptons and require that the charge values on leptons and hadrons given by the appropriate neutral current be Q_l and Q_h,

$$Q_l = T^3 - \sqrt{3}\,T^8, \quad Q_h = T^3 + \frac{1}{\sqrt{3}}\,T^8 \tag{23}$$

This gives a unique value for $\theta = 15°$ which, it so happens, is the right numerical value for Cabibbo suppression of leptonic decays of hadrons. It is easy to check that there are no neutral strangeness-changing currents. I am not presenting this as the solution for the hadron problem, firstly because in such a theory we would get a $V+A$ charged strangeness-changing quark weak current; secondly, to get any non-leptonic strangeness-changing decays at all we shall need to use a further spontaneous symmetry breaking which must employ something like a non-zero expectation value for the physical K_0-particles ($\langle K_{01} \rangle \neq 0$) in addition to the intermediate boson theory above. The hadron problem is still unsolved.

References

1. J. C. Ward & A. Salam, Nuovo Cimento *11*, 568 (1959); Phys. Lett. *13*, 168 (1964); Phys. Rev. *136*, 763 (1964).
2. A. Salam, Phys. Rev. *127*, 331 (1962).
3. S. Weinberg, Phys. Rev. Lett. *19*, 1264 (1967).

Discussion

Pais

Suppose you have by some mechanism or other a range for the effective mass of the charged W. Suppose I calculate some leptonic process, in which more than a single W-meson goes across between leptons. Is it not seriously divergent?

Salam

My claim is that in the theory I discussed there is no $k_\mu k_\nu$ term in the propagator.

Pais

In that case I want to ask a second question namely the following one. You have auxiliary scalar fields. I thought what you said is the following. The purpose of this field is by some self-consistency condition to generate the mass. At the same time you have no Goldstone particles. Suppose I have now retransformed the theory into this form. In what way does that theory now present itself to me in a different fashion from a theory in which I had an ordinary mass from the very beginning?

Salam

Formally none. Except that this mass does not have the same origin. All your troubles are now computing this mass.

Sudarshan
I would like to pursue this point a little further. Suppose, after the theory has reached its final stage, one attempts to write down rules for computing diagrams. Would I then simply substitute the $g_{\mu\nu}/(k^2-m^2)$ propagator?

Salam
Let us have it clear. There are two distinct formulations of the theory; the fields in the two formulations are being transformed one into the other by a non-linear (gauge) transformation. One is the perfect theory which contains all of the scalar particles; the gauge is preserved, etc. In that version of the formulation there is no $k_\mu k_\nu$ term and there are no divergences. If you ask me for a dictionary to the conventional theory obtained by transforming to the new set of fields where not all the scalar particles are present I haven't got this dictionary worked out yet. I conjecture that this will be a very simple dictionary but I have not evolved it yet.

Pais
So you have not proved that the $k_\mu k_\nu$ term can be dropped?

Salam
It can clearly be dropped in the symmetrical version of the theory. I believe this will also happen in the normal theory because I know there are no divergences; and also I know the two theories are equivalent so far as the final results are concerned.

Sudarshan
But in that case if I look at the propagator then the effective propagator for this particular particle does it not have a scalar negative norm particle because of the $g_{\mu\nu}$?

Salam
No. It cannot have negative norms because the symmetrical version of the theory does not have them. A local transformation of fields cannot introduce negative norms if there are none before.

Stech
I have some difficulty in understanding such a high mass for the intermediate boson. If one wants to calculate, in a very rough way, an absolute rate for non-leptonic decays as Glashow, Weinberg and Schnitzer did some time ago or when one wants to compare the electromagnetic mass differences with non-leptonic matrix elements of the current–current type one sees that a high mass for the intermediate particle is not likely. There are several arguments against a high value of a mass (Λ) representing the structure of weak interaction. My

own one is based on a comparison of electromagnetic mass differences with one-particle matrix elements of the weak interaction describing non-leptonic decay processes.

One has for an example:

$$\int \frac{d^4q}{-q^2} \int d^4x\, e^{iqx} \langle n | T(V^{\mu}_{1-i2}(x) V^{8}_{\mu}(0)) | p \rangle = \frac{\sqrt{3}\, m_p (m_n - m_p)}{e^2/4\pi}$$

and

$$\int \frac{d^4q}{-q^2+\Lambda^2} \frac{\Lambda^2}{m_p^2} \int d^4x \left\langle n \left| T(V^{\mu}_{1-i2}(x)\, V^{4-i5}_{\mu}(0) + A^{\mu}_{1+i2}(x)\, A^{4-i5}_{\mu}(0)) \right| \Lambda^0 + \frac{1}{\sqrt{3}} \Sigma^0 \right\rangle$$

$$= \frac{(2\pi)^4 \left\langle n \left| \mathcal{H}_{\text{weak}} \right| \Lambda^0 + \frac{1}{\sqrt{3}} \Sigma^0 \right\rangle}{\sin\theta \cos\theta G m_p^2/\sqrt{2}}$$

A comparison of the two integrals can be made using the assumption that the high momentum transfer region (Björken limit) decisively determines the (dominant) octet parts of the two interactions. Defining a factor x by $x = \langle n | VV + AA | \Lambda^0 + (1/\sqrt{3})\Sigma^0 \rangle / 2\sqrt{2} \langle n | VV | p \rangle$ one gets from the known numerical values of the right hand sides of the above equations (using octet parts only) $|x| \approx 0.16\, m_p^2/\Lambda^2$. By this relation high values of Λ are excluded even if a sizeable suppression of the weak interaction matrix element (in breaking $SU(3) \times SU(3)$) occurs. This conclusion is, of course, not valid if the high momentum transfer region is unimportant for the value of the integrals. In this case $|x| \approx 0.16 \overline{(1/q^2)} m_p^2$ i.e. rather low values of q^2 determine the integrals. The Björken limit would be irrelevant and could not be held responsible for the octet dominance observed in both interactions.

Salam (remark added after discussion)
The two integrals can be compared only if one neglects Λ^2 in the denominator $1/(-q^2+\Lambda^2)$. Clearly for $\Lambda^2 \approx (50\, m_p)^2$ this is questionable. So the argument based on neglecting Λ^2 in the denominator to prove that Λ^2 is small is somewhat circular.

Reprinted from ANNALS OF PHYSICS, Vol. 53, No. 1, June, 1969
Copyright © 1969 Academic Press Inc.

Finite-Component Field Representations of the Conformal Group

G. MACK*

International Centre for Theoretical Physics, Trieste, Italy

AND

ABDUS SALAM**

International Centre for Theoretical Physics, Trieste, Italy

We review work done on realization of broken symmetry under the conformal group of space-time in the framework of finite-component field theory. Topics discussed include: Most general transformation law of fields over Minkowski space. Consistent formulation of an orderly broken conformal symmetry in the framework of Lagrangian field theory; algebra of currents and their divergences; Manifestly conformally covariant fields and their couplings.

I. INTRODUCTION

The conformal symmetry of space time as a possible generalization of Poincaré symmetry has provided a recurrent theme in particle physics.[1] The problems associated with conformal symmetry are (i) its physical interpretation and (ii) the problems arising from its broken character and the precise manner of descent to Poincaré invariance.

In this paper we wish to concentrate on (ii) and present work done on realization of conformal symmetry—and in particular of the algebra associated with the group—using field operators which satisfy Lagrangian equations of motion. The fields may be defined over Minkowski space-time manifold x_μ or over a projective six-dimensional manifold η_A related to x_μ. We believe this approach to conformal symmetry offers the best hope of exploiting the symmetry physically in contrast to approaches based on a group theoretic treatment of state vector spaces associated with the group.[2] This is essentially because in the latter approach it would be much more difficult to see how to break the symmetry down to Poincaré invariance.

* Permanent address: University of Munich, Fed. Rep. Germany.

** On leave of absence from Imperial College, London, England.

[1] A historical survey may be found in Refs. (*1–3*).

[2] Unitary irreducible representations of $SU(2, 2)$ ($\approx O(2, 4)$) have been constructed in Refs. (*4–10*). The problem of reduction with respect to the Poincaré subgroup has been considered in Refs. (*11–15*) for few special cases.

The plan of the paper is as follows: In Section II we give the most general transformation law of fields (defined over Minkowski space x_μ) for conformal symmetry, using the theory of induced representations.

In Section III we enumerate the Lagrangians (for particles of spin $\leqslant 1$) which are conformally invariant and describe some modes of symmetry breaking—in particular the physically interesting case of conformal symmetry breaking to the extent of breaking dilatation invariance only. A current algebra formulation of broken dilatation symmetry has recently been found useful and gave rise to sucessful predictions for multimeson production (*16*, *17*). Here we extend this scheme to the full conformal group, considering it as a broken symmetry in the same sense as chiral $SU(3) \otimes SU(3)$. For an explicit Lagrangian model see Appendix.

Even where the formalism of Sections II and III is conformally covariant, it is not manifestly so. In Section IV we treat the manifestly covariant formulation of wave equations for quantized fields defined over a six-dimensional projective space.

In Section V we review an attempt to understand $V - A$ or $V + A$ weak interaction theory as a conformally invariant theory of fundamental interactions. Not considered in this paper are representations of the conformal group which give rise to infinite component fields. They will be discussed elsewhere (*18*).

The reader who is mainly interested in a discussion of possible physical interpretations of the conformal group is referred to Refs. (*19–25*). Mention should also be made of work of Ref. (*26*) which is somewhat different in spirit.

Without further discussion of the usefulness and consistency of possible different physical interpretations, we adopt for the purpose of the present paper the following point of view, following essentially Kastrup (*2*, *24*)

1. Exact dilatation symmetry would imply that to every physical system in any given state *sub specie aeternitatis* and for arbitrary $\rho > 0$ another one exists which is realizable in nature and differs from the first one only in that every physical observable is changed by a factor ρ^l; where l is the dimension of length of the observable in question.

2. Special conformal transformations may be interpreted as space–time dependent dilatations.

3. We consider only infinitesimal transformations. Then no problem arises with causality; in particular the local equal time commutation relations (C.R.) of fields will be invariant. (Section III.)

II. TRANSFORMATION LAW OF FIELDS

The conformal group of space time consists of coordinate transformations as follows:

1. Dilatations

$$x'_\mu = \rho x_\mu , \qquad \rho > 0$$

2. Special conformal transformations

$$x'_\mu = \sigma^{-1}(x)(x_\mu - c_\mu x^2) \tag{II.1}$$

where

$$\sigma(x) = 1 - 2cx + c^2x^2$$

3. Inhomogeneous Lorentz transformations

The generators D of dilatations, K_μ of special conformal transformations, and P_μ, $M_{\mu\nu}$ of the Poincaré group admit of the following commutation relations (C.R.):

$$\begin{aligned} [D, P_\mu] &= -iP_\mu & [D, M_{\mu\nu}] &= 0 \\ [D, K_\mu] &= iK_\mu & [K_\mu , K_\nu] &= 0 \\ [K_\mu , P_\nu] &= -2i(g_{\mu\nu}D + M_{\mu\nu}) & [K_\rho , M_{\mu\nu}] &= i(g_{\rho\mu}K_\nu - g_{\rho\nu}K_\mu) \end{aligned} \tag{II.2}$$

plus those of the Poincaré algebra. Parity must satisfy

$$\Pi D \Pi^{-1} = D, \qquad \Pi K_\mu \Pi^{-1} = \pm K_\mu , \qquad \Pi P_\mu \Pi^{-1} = \pm P_\mu \tag{II.3}$$

where the + sign stands for $\mu = 0$, and the − sign for $\mu = 1, 2, 3$.

The C.R. (II.2) can be brought into a form which exhibits the $O(2, 4)$ structure of the conformal group explicitly by defining, for $\mu, \nu = 0 \cdots 3$,

$$J_{\mu\nu} = M_{\mu\nu} , \qquad J_{65} = D, \qquad J_{5\mu} = \tfrac{1}{2}(P_\mu - K_\mu), \qquad J_{6\mu} = \tfrac{1}{2}(P_\mu + K_\mu).$$

Then

$$[J_{KL} , J_{MN}] = i(g_{KN}J_{LM} + g_{LM}J_{KN} - g_{KM}J_{LN} - g_{LN}J_{KM}) \tag{II.4}$$

where

$$g_{AA} = (+---, -+), \qquad A = 0, 1, 2, 3, 5, 6.$$

Note that the special conformal transformations do not take momentum eigenstates into momentum eigenstates, as $[K_\mu , P_\nu]$ does not commute with the momenta P_ρ. We also notice the relation (*27*)

$$e^{i\alpha D} p^2 e^{-i\alpha D} = e^{2\alpha} p^2.$$

Because of this relation, exact dilatation symmetry (with an integrable generator D that takes one-particle states into one-particle states) *implies that the mass spectrum*

is either continuous or all masses are zero. This clearly implies that exact dilatation symmetry is physically unacceptable and one will therefore have to make assumptions on the dynamics which specify how the conformal symmetry is broken. A theory of this type, which is in a sense analogous to the $SU(3) \otimes SU(3)$ current algebra with PCAC will be presented in the next section.

First we have to define, however, what we mean by an (infinitesimal) dilatation or special conformal transformation, as we want it to transform a physical system into another one that is realizable in nature (and not, e.g., a proton with mass m into some nonexistent particle with mass $\rho^{-1}m$, for arbitrary $\rho > 0$).

To do this we postulate that there exist interpolating fields to every particle which transform according to a representation of the conformal algebra, i.e.

$$(T(g)\,\varphi)_\alpha(x) = S_{\alpha\beta}(g, x)\,\varphi_\beta(g^{-1}x) \qquad \text{for infinitesimal} \quad g \in o\,(2, 4) \quad \text{(II.5)}$$

where g acts on the coordinates x as indicated in Eq. (II.1).

It follows from Eq. (II.5) and the multiplication law for the representation matrices $T(g)$ that

$S(g, 0)$ *must be a representation of the stability subgroup of* $x = 0$.

It is seen from Eq. (II.1) that this subgroup (the little group in physical usage) which leaves $x = 0$ invariant is given by special conformal transformations, dilatations and homogeneous Lorentz transformations. From the C.R. Eq. (II.2) one finds that the Lie algebra of this subgroup is isomorphic to a Poincaré algebra + dilatations, i.e. we have

$$(SO(3,1) \otimes \{D\}) \circledS T_4\,. \qquad \text{(II.6)}$$

The 4-dimensional translation subgroup T_4 corresponds to the special conformal transformations, and $SO(3,1)$ is the spin part of the Lorentz group.

Given any representation $S(g, 0)$ of the little group (II.6) we can now determine, in accordance with the standard theory of induced representations (*28*), the complete action of the generators of the conformal group on the field $\varphi(x)$ as follows:

Let $\Sigma_{\mu\nu}$, Δ, κ_μ be the infinitesimal generators of the little group (II.6) corresponding to Lorentz transformations, dilatations, and special conformal transformations, respectively. They satisfy

$$\begin{aligned} &[\kappa_\mu, \kappa_\nu] = 0, \qquad [\Delta, \kappa_\mu] = +i\kappa_\mu \\ &[\kappa_\rho, \Sigma_{\mu\nu}] = i(g_{\rho\mu}\kappa_\nu - g_{\rho\nu}\kappa_\mu), \\ &[\Sigma_{\rho\sigma}, \Sigma_{\mu\nu}] = i(g_{\sigma\mu}\Sigma_{\rho\nu} - g_{\rho\mu}\Sigma_{\sigma\nu} - g_{\sigma\nu}\Sigma_{\rho\mu} + g_{\rho\nu}\Sigma_{\sigma\mu}) \end{aligned} \qquad \text{(II.7)}$$

Choose the basis in index space in such a way that space time translations do not act on the indices, i.e.

$$P_\mu \varphi_\alpha(x) = i \frac{\partial}{\partial x^\mu} \varphi_\alpha(x) \tag{II.8}$$

It follows that for every element X of the conformal algebra

$$X\varphi(x) = \exp(-iP_\mu x^\mu)\, X' \varphi(0)$$

where

$$\begin{aligned} X' &= \exp(+iP_\mu x^\mu)\, X \exp(-iP_\mu x^\mu) \\ &= \sum_{n=0}^{\infty} \frac{i^n}{n!} x^{\nu_1} \cdots x^{\nu_n} [P_{\nu_1}, [\cdots [P_{\nu_n}, X] \cdots]]. \end{aligned} \tag{II.9}$$

The important point is that the sum on the *RHS.* of Eq. (II.9) is actually finite.[3] From the C.R. Eq. (II.2) it is found by inspection that there are at most three non-vanishing terms in this sum. Evaluating the finite multiple commutators, e.g., for $X = K_\mu$, we get

$$\exp(+iP_\nu x^\nu) K_\mu \exp(-iP_\nu x^\nu) = K_\mu - 2x^\nu (g_{\mu\nu} D + M_{\mu\nu}) + 2x_\mu x^\nu P_\nu - x^2 P_\mu .$$

From this we now deduce the action of K_μ, D, $M_{\mu\nu}$ on $\varphi(x)$, since the action on $\varphi(0)$ is known by hypothesis; e.g. $K_\mu \varphi(0) = \kappa_\mu \varphi(0)$. The final results are

$$\begin{aligned} P_\mu \varphi(x) &= i\partial_\mu \varphi(x) \\ M_{\mu\nu} \varphi(x) &= \{i(x_\mu \partial_\nu - x_\nu \partial_\mu) + \Sigma_{\mu\nu}\}\, \varphi(x) \\ D\varphi(x) &= \{ix_\nu \partial^\nu + \Delta\}\, \varphi(x) \\ K_\mu \varphi(x) &= \{i(2x_\mu x_\nu \partial^\nu - x^2 \partial_\mu - 2ix^\nu [g_{\mu\nu}\Delta + \Sigma_{\mu\nu}]) + \kappa_\mu\}\, \varphi(x) \end{aligned} \tag{II.10}$$

where the matrices $\Sigma_{\mu\nu}$, Δ, κ_μ satisfy the C.R. (II.7).

We have thus shown that all field theoretically admissible representations of the conformal algebra are induced by a representation of the algebra of the little group (II.6). Since this algebra has two nontrivial ideals (=invariant subalgebras) $\{D\} \otimes T_4 \supset T_4$ there arise the following types of representations:

I. finite-dimensional representations of the little group

(a) $\kappa_\mu = 0$

(b) $\kappa_\mu \neq 0$ but nilpotent

[3] The finiteness of this sum follows also from a general theorem of O'Raifertaigh's (*29*): Let G be any finite dimensional Lie algebra containing the Poincaré algebra $P = SO(3,1) \otimes T$. Then there exists a finite number n_0 such that

$$[T, [T, [\ldots T, X] \ldots]]] = 0 \text{ for } n > n_0 \text{ and any } X \in \mathrm{G}.$$

II. infinite-dimensional representations of the little group.
Regarding these representations and their physical uses, the following remarks are in order:

(1) For case (Ia) $\Delta = il\mathbf{1}$ by Schur's lemma if the $\Sigma_{\mu\nu}$ form an irreducible representation of the homogeneous Lorentz algebra.

(2) In Section III it will be shown that the notion of a (broken) conformal symmetry admits of a perfectly consistent formulation in the framework of ordinary Lagrangian field theory. This theory makes use of finite-dimensional representations of the little group (II.6), with $\kappa_\mu = 0$ (type Ia). All generators will be hermitian.

(3) For case (Ib) the conclusion that all the κ_μ must be nilpotent follows from the well-known fact that in any finite-dimensional representation of the Poincaré algebra the generators of translations are nilpotent (*30*).[4]

(4) The possibility of using representations of type (Ib) for physical purposes could be interesting because it may give rise to spin multiplets. The problem is that the representations induced in this way are apparently not fully reducible.
Further discussion on this possibility will be given in Section V, where Hepner's work on the use of these representations will be reviewed (*31*).

(5) The possible use of infinite-dimensional representations of the little group will not be discussed in this paper. This would lead to the consideration of infinite-component field theories which will be discussed elsewhere (*18*).

III. LAGRANGIAN FIELD THEORY; APPLICATION TO STRONG AND ELECTROMAGNETIC INTERACTION

In the present section we shall show that the idea of an orderly broken symmetry under the conformal group of space time admits of a perfectly consistent formulation in the framework of ordinary Lagrangian field theory. The considerations presented here are an extension of an unpublished note by one of the present authors (*32*). For simplicity we assume fields with spin $\leqslant 1$ and minimal couplings.[5] We shall show that:

(1) There exist local conformal currents $\mathscr{K}_{\nu\mu}$ and a dilatation current $\mathscr{D}_\mu$ such that the corresponding generators K_μ and D are hermitian and have C.R. with the particle fields as given in Eq. (II.10), with $\Delta = il\mathbf{1}$, $\kappa_\mu = 0$ (type Ia). *This is true independently of whether the action* $\int \mathscr{L}\, d^4x$ *is invariant or not.*

[4] Recall that nilpotency of a matrix κ means that $\kappa^m = 0$ for a suitable positive integer m.
[5] Minimal couplings are those involving no derivatives of fermion fields and up to first order derivatives of boson fields; they have the virtue of not affecting the canonical C.R. of the fields.

(2) The kinetic energy term without mass is conformal invariant.[6] The same is true of all non-derivative couplings with dimensionless coupling constants and all couplings arising from (Yang–Mills type) (*34*) gauge field theories. This includes electromagnetism. It also includes weak interactions mediated by an intermediate boson, if this boson is associated with a gauge field associated with some internal group (e.g., for hadrons, the Cabbibo $SU(2)$ subgroup of one of the $SU(3)$ ideals of chiral $SU(3) \otimes SU(3)$, or the $U(2) \otimes U(2)$ group considered for leptons by Ward and Salam (*36*), which includes both *EM* and weak interactions.)

(3) Besides the exact symmetry limit corresponding to massless particles only, the possibility also exists of a spontaneous breakdown of conformal symmetry. There, all particles can be massive except for $I = 0$, $J^P = 0^+$ massless Goldstone boson. An example of a corresponding Lagrangian (the σ model of Gell-Mann and Lévy) is discussed in the Appendix.

Abstracting from Lagrangian field theory, a current algebra type scheme may be set up. It is composed of the C.R. of the currents with the particle fields, Eqs. (III.1) and (III.6), and the relation (III.17) between the divergences of the currents. In addition, an algebraic relation between the divergence of the dilatation current and the axial vector currents of chiral $SU(3) \otimes SU(3)$ has been proposed elsewhere (*16*) and is given in Eqs. (A.4) and (A.5) of the Appendix. Equation (III.17) expresses the idea that the breaking of conformal symmetry is minimal in the sense that there is only as much breaking of the conformal symmetry as is induced by the breaking of dilatation symmetry alone.

1. *The conformal currents*

According to Eq. (II.10) we want to transform the interpolating fields as follows

$$\begin{aligned} &\text{(a)} \quad D\varphi(x) = i(-l + x_\nu \partial^\nu)\, \varphi(x) \\ &\text{(b)} \quad K_\mu \varphi(x) = i(-2lx_\mu + 2x_\mu x_\nu \partial^\nu - x^2 \partial_\mu - 2ix^\nu \Sigma_{\mu\nu})\, \varphi(x) \\ &\text{(c)} \quad M_{\mu\nu}\varphi(x) = i(x_\mu \partial_\nu - x_\nu \partial_\mu - i\Sigma_{\mu\nu})\, \varphi(x) \end{aligned} \tag{III.1}$$

If φ is the electromagnetic vector potential, we may postulate this transformation law only up to a gauge transformation. Eq. (III.1) is to be understood in this sense in the following. The well known (*35*), (*27*) reason for this is that, for a massless

[6] The conformal invariance of the free Maxwell equations and the free massless Klein–Gordon, Dirac- and Weyl- equations has been known for a long time, cf. Refs. (*1–3*).

A conformal invariant spinor equation with a nonlinear interaction $(\bar{\psi}\psi)^{4/3}$ has been considered by Gürsey (*1*)

Conformal invariant wave equations in Minkowski space for massless free particles with arbitrary spin have been considered in refs. (*12*, *15*, and *33*).

particle, the vector potential is not a manifestly Lorentz covariant field in the ordinary sense. Through the last equation, $\Sigma_{\mu\nu}$ is defined in terms of the spin of the particle:

$$\begin{aligned}
&\text{when acting on a spin 0 field} && \Sigma_{\mu\nu} = 0\\
&\text{when acting on a spin } \tfrac{1}{2} \text{ field} && \Sigma_{\mu\nu} = \frac{i}{4}[\gamma_\mu, \gamma_\nu]\\
&\text{when acting on a vector field} && (\Sigma_{\mu\nu}A)_\rho = i(g_{\mu\rho}A_\nu - g_{\nu\rho}A_\mu)
\end{aligned}$$

The present theory does not lead to multiplets of particles with different spin. We fix the values of l to be

$$\begin{aligned}
l &= -1 \quad \text{for sclar and vector fields}\\
l &= -\tfrac{3}{2} \quad \text{for spin } \tfrac{1}{2} \text{ fields}
\end{aligned} \tag{III.2}$$

and $l = -1 - j$ for $(0, j) \oplus (j, 0)$ fields for arbitrary spin j, cf. ref. *15*). This choice is necessary in order to obtain acceptable currents because only then the canonical equal time commutation relations of the fields are invariant.[7] The values of l in (III.2) agree with the actual dimension of length of the fields in question. Note that (III.1c) is of the form $\varphi'(x) = \rho^l\varphi(\rho^{-1}x)$ under $x_\mu \to \rho x_\mu$; $\rho = 1 + \epsilon$ so that the fields transform under dilatations according to their actual dimension of length.

We can now write down the following local currents:

$$\begin{aligned}
\mathscr{D}_\nu(x) &= x^\rho T_{\nu\rho} - \sum_r l_r \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi_r}\varphi_r\\
\mathscr{K}_{\nu\mu}(x) &= 2x^\rho x_\mu T_{\nu\rho} - x^2 T_{\nu\mu} - \sum_r \left(\frac{\partial \mathscr{L}}{\partial \partial^\nu \varphi_r}(2l_r x_\mu + 2ix^\rho \Sigma_{\mu\rho})\,\varphi_r\right) - \sum_{\substack{\text{spin 0}\\ \text{fields}}} g_{\nu\mu}\phi^+\phi
\end{aligned} \tag{III.3}$$

For the free field case with exact symmetry, similar currents have been written down by McLennan (*33*) and Wess (*27*).

The angular momentum current has the usual form. The energy momentum tensor is defined by

$$T_{\nu\rho}(x) = -g_{\nu\rho}\mathscr{L} + \sum_r \frac{\partial \mathscr{L}}{\partial \partial^\nu \varphi_r}\partial_\rho \varphi_r$$

[7] Eq. (III.2) also follows from the requirement that the free massless wave equation be fully conformal invariant.

It is convenient to choose the kinetic energy term in the Lagrangian as:

$$\mathscr{L}_0 = \sum_{\text{spin}\frac{1}{2}} \bar{\psi}(\tfrac{1}{2}\gamma_\mu \overleftrightarrow{\partial^\mu} - M)\,\psi + \sum_{\text{spin}0} \tfrac{1}{2}(\partial_\mu \varphi^+ \partial^\mu \varphi - \mu^2 \varphi^+ \varphi)$$

$$- \sum_{\text{spin}1} \{\tfrac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 + \tfrac{1}{2}m^2 B_\nu B^\nu\} \qquad \text{(III.4)}$$

Since the zeroth components of the currents, $T_{0\mu}$, $\mathscr{D}_0$, $\mathscr{K}_{0\mu}$ and $\mathscr{M}_{0\nu\lambda}$ are all hermitian, the corresponding generators, formally defined as the space integrals of the zeroth component of these currents, are also hermitian.

One checks by using the canonical equal time C.R. of the fields

$$[\phi(x), \pi(x')]_\pm = i\delta^3(\mathbf{x} - \mathbf{x}') \qquad \text{for} \quad x_0 = x_0' \qquad \text{(III.5)}$$

that

$$D\varphi(x) \equiv -\int_v d^3x[\mathscr{D}_0(x'), \varphi(x)]_{x_0 = x_0'}$$
$$K_\mu\varphi(x) \equiv -\int_v d^3x[\mathscr{K}_{0\mu}(x'), \varphi(x)]_{x_0 = x_0'} \qquad \text{(III.6)}$$

satisfy Eqs. (III.1). This is independent of whether the Lagrangian is conformal invariant or not. (The integration in (III.4) goes over some volume v including $\mathbf{x} = \mathbf{x}'$.)

2. *Divergence of the Currents.*

We now turn to the discussion of the properties of the divergence of the dilatation current $\mathscr{D}_\nu$ and special conformal currents $\mathscr{K}_{\nu\mu}$. The dynamical information of a symmetry (exact or broken) defined by the transformation law of the fields, lies in the properties of the divergence of the corresponding currents.

From Eq. (III.3) it is seen that the dilatation current $\mathscr{D}_\nu(x)$ depends on x explicitly and can therefore not be coupled to the field of a vector particle. However, using energy momentum conservation $\partial^\nu T_{\nu\rho} = 0$, we find for its divergence

$$\partial^\nu \mathscr{D}_\nu(x) = T_\nu^{\ \nu} - \partial_\nu \left\{\sum_r l_r \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi_r} \varphi_r\right\} \qquad \text{(III.7)}$$

From this we see that:

The divergence of the dilatation current is a local
$I = 0$, $J^P = 0^+$ *field.*

Using the Euler Lagrange equation of motion, (III.7) can be rewritten in the form (*37*)

$$\partial^\nu \mathscr{D}_\nu(x) = -4\mathscr{L} - \sum_r \left(l_r \frac{\partial \mathscr{L}}{\partial \varphi_r} \varphi_r + (l_r - 1) \frac{\partial \mathscr{L}}{\partial \partial_\mu \varphi_r} \partial_\mu \varphi_r \right) \tag{III.8}$$

From Eq. (III.8) it is seen that all pieces of the Lagrangian that do not involve constants with nonzero dimension of length give zero contribution. To see this, notice that the RHS of Eq. (III.8) vanishes, since it is simply the Euler equation for a homogenous function (all terms having the same dimensionality). This establishes that *the kinetic energy term without mass is dilatation invariant, and so are all couplings with a dimensionless coupling constant.*

Next we turn to the divergence of the conformal currents. Using energy momentum conservation we find from Eq. (III.3)

$$\begin{aligned} \partial^\nu \mathscr{K}_{\nu\mu}(x) = 2x_\mu & \left(T^\nu{}_\nu - \partial_\nu \left\{ \sum_r l_r \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi_r} \varphi_r \right\} \right) \\ & - \sum_r \left\{ 2l_r \frac{\partial \mathscr{L}}{\partial \partial^\mu \varphi_r} \varphi_r + 2i \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi_r} \Sigma_{\mu\nu} \varphi_r \right\} + \sum_{\text{spin 0}} \partial_\mu (\varphi^+ \varphi) \\ & + 2x^\rho \left\{ T_{\mu\rho} - T_{\rho\mu} + 2i\, \partial_\nu \left(\sum_r \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi_r} \Sigma_{\mu\rho} \varphi_r \right) \right\} \end{aligned} \tag{III.9}$$

The last line is equal to $2x^\rho \partial^\nu \mathscr{M}_{\nu\mu\rho}$ and vanishes by angular momentum conservation. Hence

$$\partial^\nu \mathscr{K}_{\nu\mu} = 2x_\mu \partial^\nu \mathscr{D}_\nu + V_\mu \tag{III.10}$$

where V_μ is a local vector field defined by the second line on the RHS of Eq. (III.9). From Eq. (III.4) one checks by explicit calculation that the kinetic energy term without mass gives zero contribution to V_μ. Since it gives no contribution to $\partial^\nu \mathscr{D}_\nu$ either, we see that the kinetic energy term without mass is fully conformal invariant. Furthermore, from (III.10) the condition that an interaction Lagrangean $\mathscr{L}_I$ be fully conformal invariant is found to be the following

(1) It is dilatation invariant, i.e., has a dimensionless coupling constant

$$(2) \quad \sum_r \left(l_r \frac{\partial \mathscr{L}_I}{\partial \partial^\mu \varphi_r} \varphi_r + i \frac{\partial \mathscr{L}_I}{\partial \partial_\nu \varphi_r} \Sigma_{\mu\nu} \varphi_r \right) = 0; \quad \mu = 0 \cdots 3 \tag{III.11}$$

Condition (2) is independent from condition (1) as is seen from the example $\mathscr{L}_I = gA^\mu \pi (\partial_\mu \sigma)$ which satisfies (1) but not (2). An example which satisfies neither (1) nor (2) is the derivative pion nucleon coupling $\bar{N} \gamma_5 \gamma_\mu N \partial^\mu \pi$. For nonderivative couplings, condition (2) is trivially satisfied.

3. *Yang–Mills theory*

Let us now turn to the question of the conformal invariance of the coupling of vector gauge fields in a Yang–Mills type gauge field theory. Let A be the internal n-parameter symmetry algebra, and $B_\mu{}^a$, $a = 1 \cdots n$, the corresponding gauge vector fields. Under an infinitesimal transformation with constant infinitesimal parameter ϵ^a, all fields transform according to

$$\Phi'_A = \Phi_A + \delta\Phi_A, \qquad \delta\Phi_A = i\epsilon^a T^B_{a,A}\Phi_B \tag{III.12}$$

where the matrices $\mathbf{T}_a$ form a hermitian representation of the algebra **A**. Hermiticity reads

$$(T^B_{a,A})^* = T^A_{a,B} \tag{III.13}$$

As is well known (*34*, *38*), all couplings of the vector fields $B_\mu{}^a$ are completely determined from the postulates of a Yang–Mills type theory and are obtained by the substitution

$$\partial_\mu\Phi_A \to \partial_\mu\Phi_A - igT^B_{a,A}B_\mu{}^a\Phi_B \tag{III.14}$$

Here, g is a dimensionless real coupling constant. To test for conformal invariance we see that condition (1) above is always fulfilled, while condition (2) is also satisfied because the only derivative couplings are the couplings of mesons, which have the form

$$\mathscr{L}_I = -\frac{i}{2}g\{(\partial^\mu\Phi^A)^*(T^B_{a,A})^*\Phi_B - (\partial^\mu\Phi^A)\,T^B_{a,A}\Phi^*_B\}\,B_\mu{}^a \tag{III.15}$$

for spin 0 fields Φ_A, and

$$\mathscr{L}_I = \frac{i}{2}g\{(\partial^\mu\varphi^{\nu A})^*(T^B_{a,A})^*(\varphi_{\nu B}B_\mu{}^a - \varphi_{\mu B}B_\nu{}^a) - h.c.\} \tag{III.16}$$

for spin 1 fields $\varphi_A{}^\nu$. Inserting into Eq. (III.11) and making use of Eqs. (III.13) and (III.2) one finds that condition (2) is indeed satisfied.

It is now tempting to speculate that in physics there is only as much breaking of conformal symmetry as is induced by the breaking of dilatation symmetry. In other words there should be a remainder of conformal symmetry in the sense that all couplings satisfy condition (2) above. This is equivalent to the algebraic condition

$$\partial^\nu\mathscr{K}_{\nu\mu} = 2x_\mu\partial^\nu\mathscr{D}_\nu \tag{III.17}$$

The virtue of this restriction is that it still allows for a breaking of the symmetry by the mass terms in the Lagrangian.

IV. MANIFESTLY $O(2, 4)$ COVARIANT FIELD TRANSFORMATION LAW

General experience from the history of elementary particle physics may lead one to the opinion that "the only good covariance is a manifest covariance". This is the motivation for the present section. It is mainly pedagogical in character and much of the material presented may be found in the literature for special cases, but is presented here in a unified way. Manifestly conformal invariant free field equations were first discussed by Dirac (*39*). Invariant interactions were given by Kastrup (*40*).

The problems to be solved are the following:

(1) Write down manifestly conformal invariant transformation laws for fields.

(2) Determine the relation between the old fields $\varphi_\alpha(x)$ occurring in the transformation law Eq. (II.10) and the new fields which are transformed manifestly covariantly.

(3) Write down manifestly invariant free field equations and interactions.

The new fields will be multispinor functions on the four-dimensional surface in a five-dimensional projective space rather than Minkowski space. Their physical interpretation will nevertheless be guaranteed by correspondence with ordinary fields $\varphi_\alpha(x)$ over Minkowski space discussed in Section II. This correspondence also allows one to consider questions of unitarity and quantization by reference to Minkowski space.

1. *Manifestly Covariant Transformation Law for Fields*

A manifestly $O(2, 4)$ covariant transformation law may be written down for multispinor functions $\chi_\beta(\eta)$ defined on the five-dimensional hypersurfaces of R^6 given by

$$\eta_B \eta^B = L^2 \tag{IV.1}$$

and satisfying $\chi_\beta(\eta) = \chi_\beta(-\eta)$.

Summation over B is over 0, 1, 2, 3, 5, 6 with metric $(+ \; - - - ; \; - +)$. There are three essentially different surfaces, corresponding to $L^2 = \pm 1, 0$.

Suppose that s_{AB} is any finite dimensional representation of the algebra of $O(2, 4)$ ($\approx SU(2, 2)$) acting on the indices of $\chi_\beta(\eta)$ only. Then a manifestly covariant transformation law including an orbital part is given by (cf. Eq. (II.4))

$$\delta\chi(\eta) \equiv -i\epsilon^{AB} J_{AB} \chi(\eta) = -i\epsilon^{AB}(L_{AB} + s_{AB})\, \chi(\eta) \tag{IV.2}$$

where

$$L_{AB} = i(\eta_A \partial_B - \eta_B \partial_A), \qquad \partial_B \equiv \frac{\partial}{\partial \eta^B}$$

Clearly L_{AB} and s_{AB} commute with each other and satisfy the C.R. of $O(2, 4)$ separately. It is important to notice that L_{AB} is a well defined operator when acting on functions that are only defined on the hypersurface (IV.1) (*39*). This is so because L_{AB} corresponds to an infinitesimal coordinate transformation which is a pseudo-rotation of the hypersurface (V.1) into itself.

The cone

$$\eta_B \eta^B = 0 \tag{IV.3}$$

is also left invariant by the coordinate transformation $\eta_B \to \lambda \eta_B$, $\lambda > 0$. Moreover, this transformation commutes with the $O(2, 4)$ rotations. Therefore we may require the fields to be homogeneous functions on the cone (IV.3)

$$\chi(\lambda\eta) = \lambda^n \chi(\eta), \qquad \text{i.e.} \quad \eta^B \partial_B \chi(\eta) = n\chi(\eta). \tag{IV.4}$$

These homogeneous functions then depend arbitrarily only on 4 of the 5 coordinates which determine a point on the cone (V.3), i.e., just as many as there are coordinates in Minkowski space.[8] We shall restrict our attention to this case in the following.

2. *Mathematical preliminaries.*

Before proceeding, we need to know some mathematical lemmas.

LEMMA 1. *By a suitable choice of basis, each finite dimensional representation of the algebra* (II.7) *by matrices* $\Sigma_{\mu\nu}$, Δ, κ_μ *can be brought into the form*

$$\Sigma_{\mu\nu} = \begin{pmatrix} * & & & 0 \\ & * & & \\ & & \ddots & \\ 0 & & & * \end{pmatrix} \quad \text{(a)} \qquad \Delta = -i \begin{pmatrix} \lambda_0 1 & & & \\ & \lambda_1 1 & & \\ & & \ddots & \\ & & & \lambda_m 1 \end{pmatrix} \quad \text{(b)}$$

$$\kappa_\mu = \begin{pmatrix} 0 & & & 0 \\ & 0 & & \\ & & \ddots & \\ * & & & 0 \end{pmatrix} \quad \text{(c)} \tag{IV.5}$$

Note in particular that all κ_μ *are nilpotent, as* $(\kappa_\mu)^m = 0$, *and their top rows are identically zero.*

[8] A subtle point here is that the points where at least one coordinate x_μ is infinite may also correspond to finite coordinates on the cone. This is however not important here as we consider only infinitesimal transformations.

Proof. Finite dimensional representations of $SL(2C)$ algebra are fully reducible, whence (a). As $[\Sigma_{\mu\nu}, \Delta] = 0$, (b) follows by Schur's lemma. Without loss of generality, $\operatorname{Re}\lambda_0 \geqslant \operatorname{Re}\lambda_1 \geqslant \cdots \geqslant \operatorname{Re}\lambda_m$. By C.R. (II.7), each κ_μ is a stepoperator lowering the eigenvalue of $i\Delta$ by 1. This implies (c).

We also need some properties of finite dimensional representations of $O(2, 4)$ algebra. We make use of its isomorphy with $SU(2, 2)$ algebra and the well known Weyl's unitary trick to arrive at

LEMMA 2. *All finite dimensional representations of $O(2, 4)$ algebra (without parity) by matrices $s_{AB} = -s_{BA}$ can be obtained by reducing out tensor products of the following two inequivalent fundamental 4-dimensional representation $\Delta^{(+)}$ and $\Delta^{(-)}$ given by matrices γ_{AB} as follows:*

$$\begin{aligned}
\Delta^{(+)}:\ \gamma_{\mu\nu} &= \frac{i}{4}[\gamma_\mu, \gamma_\nu]; & \gamma_{\mu 5} &= -\frac{i}{2}\gamma_\mu\gamma_5 \\
\gamma_{56} &= -\tfrac{1}{2}\gamma_5; & \gamma_{\mu 6} &= \tfrac{1}{2}\gamma_\mu \\
\Delta^{(-)}:\ \dot\gamma_{\mu\nu} &= \frac{i}{4}[\gamma_\mu, \gamma_\nu]; & \dot\gamma_{\mu 5} &= \frac{i}{2}\gamma_\mu\gamma_5 \\
\dot\gamma_{56} &= \tfrac{1}{2}\gamma_5; & \dot\gamma_{\mu 6} &= \tfrac{1}{2}\gamma_\mu
\end{aligned} \tag{IV.6}$$

γ_μ are Dirac matrices defined by $\{\gamma_\mu, \gamma_\nu\}_+ = 2g_{\mu\nu}$, and $\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3$. All matrices satisfy $A\gamma_{BC}A = \gamma^+_{BC}$, where $A = \gamma_0$ for the usual choice of Dirac matrices with γ_0 and $i\gamma_j$ hermitian.

Parity transforms $\Delta^{(+)}$ into $\Delta^{(-)}$ and vice versa. The two representations $\Delta^{(+)}$ and $\Delta^{(-)}$ are conjugate to each other.

This is a weak form of a standard result (*41*).

The simplest nontrivial representation of $O(2, 4)$ with parity is eightdimensional and unique up to a choice of basis. It is given by $\Delta = \Delta^{(+)} \oplus \Delta^{(-)}$. It is convenient to carry out a basis transformation such that

$$\begin{aligned}
\gamma_{\mu\nu} &= \frac{i}{4}[\gamma_\mu, \gamma_\nu], & \gamma_{\mu 5} &= -\frac{i}{2}\gamma_\mu\tau_2 \\
\gamma_{56} &= -\frac{i}{2}\tau_3 & \gamma_{\mu 6} &= \tfrac{1}{2}\gamma_\mu\tau_1 \\
&& \Pi &= \gamma_0
\end{aligned} \tag{IV.7}$$

Parity is represented by γ_0 here. τ_i are Pauli matrices. In this form the eight dimensional representation has been given by Murai (*42*). For this representation a Clifford algebra exists such that (*39, 40*)

$$\frac{i}{4}[\beta_A, \beta_B] = \gamma_{AB}; \qquad \{\beta_A, \beta_B\} = 2g_{AB} \tag{IV.8}$$

The matrices β_A transform as a six-vector in the sense that

$$s(\alpha)\,\beta^A s^{-1}(\alpha) = \alpha^A{}_B \beta^B, \qquad \text{i.e.} \quad [\beta_A\,, \gamma_{BC}] = i(g_{AB}\beta_C - g_{AC}\beta_B)$$

The real numbers $\alpha^A{}_B$ are the parameters characterizing an element of $O(2, 4)$ group. This transformation law will be important for constructing invariant couplings. Explicitly the β_A may be given by

$$\beta_\mu = \gamma_\mu \tau_3\,; \qquad \beta_5 = i\tau_1\,; \qquad \beta_6 = \tau_2 \tag{IV.9}$$

There is also a conformal pseudoscalar β_7

$$\beta_7 = -i\beta_0\beta_1\beta_2\beta_3\beta_5\beta_6 = i\gamma_5\tau_3 \qquad \Pi\beta_7\Pi = -\beta_7\,.$$

All matrices satisfy

$$A\gamma_{BC}A = \gamma^+_{BC} \qquad \text{and} \qquad A\beta_B A = -\beta_B{}^+, \qquad \text{where} \qquad A = \gamma_0\tau_1 \tag{IV.10}$$

We further need the following consequence of lemma 2:

LEMMA 3. *Suppose we are given a representation of $SU(2, 2)$ algebra by matrices s_{AB} acting on a finite dimensional vector space $\mathscr{H}$. Evidently this provides at the same time a representation of the subalgebra* (II.6)*f by matrices*

$$\kappa_\mu = s_{6\mu} - s_{5\mu}\,, \qquad \Sigma_{\mu\nu} = s_{\mu\nu}\,; \qquad \Delta = s_{65}\,. \tag{IV.11}$$

We assert that there exists a projection operator E such that

(1) $\kappa_\mu E \neq 0$, (IV.12)

and

(2) $\kappa'_\mu \equiv E\kappa_\mu E = 0, \qquad \Sigma'_{\mu\nu} \equiv E\Sigma_{\mu\nu}E, \qquad \Delta' \equiv E\Delta E - n\mathbf{1}$

form a new representation of the algebra of the little group (II.6) *acting on $\mathscr{H}' = E\mathscr{H}$.*

n may be any c-number.

Moreover, every representation of the algebra of the little group with $\kappa_\mu = 0$ may be obtained in this way from a suitable representation of $SU(2, 2)$ algebra.

A constructive proof will be given below.[9] In the basis of Lemma 1, $\mathscr{H}'$ is spanned by vectors of the form

$$\begin{pmatrix} * \\ \hline 0 \\ \vdots \\ 0 \end{pmatrix}.$$

[9] Note that this construction does not amount to restriction to an invariant subspace under the full little group, as obviously $[E, \kappa_\mu] \neq 0$. Instead we are here exploiting the fact that our little group is not semisimple, and its finite dimensional representations are in general not completely reducible. $\kappa_\mu E \neq 0$ will be important later for the construction of invariant wave-equations.

It follows from (IV.5) that κ_μ maps these into the orthogonal complement $\mathscr{H}'^{\perp}$ of $\mathscr{H}'$. Consequently $E\kappa_\mu E = 0$. However, none of the matrices $\Sigma_{\mu\nu}$, Δ, κ_μ makes a transition $\mathscr{H}'^{\perp} \to \mathscr{H}'$, as $\mathscr{H}'$ and $\mathscr{H}'^{\perp}$ are separately invariant under $\Sigma_{\mu\nu}$ and Δ by virtue of Lemma 1, and κ_μ has the form (IV.5). Thus $\mathscr{H}'$ forms a representation space for the algebra $\{\Sigma'_{\mu\nu}, \Delta'\}$, extended to one of (II.7) by adjoining $\kappa'_\mu = E\kappa_\mu E = 0$. This completes the demonstration of the first part of the lemma.

It is useful to reformulate in matrix language. With a choice of basis as in Lemma 1, the matrices $(\Sigma'_{\mu\nu})_{\alpha\beta}$, $(\Delta' + \text{in } \mathbf{1})_{\alpha\beta}$, $(\kappa'_\mu)_{\alpha\beta} = 0$ are submatrices of $\Sigma_{\mu\nu}$, Δ, κ_μ with $\alpha, \beta \in I_{\text{phys}}$. I_{phys} is the set of indices α such that the basisvectors $\mathbf{e}_\alpha \in \mathscr{H}'$, where

$$\mathscr{H}' = \mathscr{H} - \bigcup_\mu \kappa_\mu \mathscr{H}$$

Finally, addition of a multiple of the identity to Δ' does obviously not change the C.R.

The second part of the Lemma 3 may be proved by explicit construction.

Consider the multispinor which transforms as direct product of 4-dimensional fundamental representations $\Delta^{(+)}$ and $\Delta^{(-)}$

$$\delta\xi_{\alpha_1\ldots\alpha_j;\dot\beta_1\ldots\dot\beta_k} = -i\epsilon^{AB}\left\{\sum_{i=1}^{j}(\gamma_{AB})_{\alpha_i\alpha'_i}\xi_{\alpha_1\ldots\alpha'_i\ldots\alpha_j;\dot\beta_1\ldots\dot\beta_k}\right.$$

$$\left. + \sum_{i=1}^{k}(\dot\gamma_{AB})_{\dot\beta_i\dot\beta'_i}\xi_{\alpha_1\ldots\alpha_j;\dot\beta_1\ldots\dot\beta'_i\ldots\dot\beta_k}\right\}; \qquad \alpha_i, \beta_i = 1\cdots 4 \tag{IV.13}$$

As a necessary condition for irreducibility we may impose symmetry properties in indices of the same kind, and tracelessness

$$\delta^{\alpha'_i}_{\dot\beta_i}\xi^{\dot\beta_1\dot\beta_2\ldots}_{\alpha_1\alpha_2\ldots} = 0.$$

The multispinor with upper dotted indices is obtained by applying the index raising operator $C\gamma_0$, with $C\gamma_\mu C = -\bar\gamma_\mu$.

Since $[\Sigma'_{\mu\nu}, \Delta'] = 0$ and $\kappa'_\mu = 0$ we may restrict our attention to an irreducible representation of $SL(2C)$ which may be characterized by a pair $(j/2, k/2)$ in the usual way.

The desired representation of $SU(2, 2)$ acts on multispinors

$$\xi_{\{\alpha_1\ldots\alpha_j\};\{\dot\beta_1\ldots\dot\beta_k\}} \equiv \xi_{\{\alpha\},\{\beta\}} \tag{IV.14}$$

which are totally symmetric in each kind of indices and have j undotted and k dotted indices. Each index takes 4 values. The dot-notation has been adopted in anticipation of connection which will be made with Lorentz group.

We now adopt a basis for the Dirac matrices as follows

$$\gamma_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \qquad \gamma_i = \begin{pmatrix} 0 & -\sigma_i \\ \sigma_i & 0 \end{pmatrix}; \qquad i\gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{IV.15}$$

Then projection onto $\mathscr{H}'$ amounts to considering only those multispinor components with

$$\alpha_i = 3, 4; \qquad \dot{\beta}_i = 1, 2$$

for all undotted indices α_i and dotted indices $\dot{\beta}_i$, respectively. Using (IV.11), (IV.13), (IV.6) it is readily established that

$$\delta\xi_{\{\alpha\},\{\dot{\beta}\}} \equiv -i\epsilon^{\mu}(\kappa_\mu \xi)_{\{\alpha\},\{\dot{\beta}\}} = 0.$$

Moreover, under infinitesimal Lorentz transformations these multispinor components are found to transform precisely like the familiar Lorentz multispinor with j undotted and k dotted indices, i.e. according to the representation $(j/2, k/2)$ as desired. Finally, $\Delta' = -il\mathbf{1}$ by Schur's lemma. This completes the proof.

Remark: *$\mathscr{H}'$ may also be characterized as the eigenspace belonging to the highest eigenvalue of is_{65}.*

Provided s_{AB} is irreducible, this follows from the fact that κ_μ is a step operator which lowers the eigenvalue of is_{65} by one unit. Irreducibility is true for traceless multispinors of the form (IV.14) but will not be proved here.

We now proceed to the proof and statement of our last lemma. Suppose that we are given a field representation of the conformal algebra of the form (II.10) induced by finite dimensional representation κ_μ, $\Sigma_{\mu\nu}$, Δ of the algebra of the little group (II.6). As we have seen, there arise in this way two types of representations, $\kappa_\mu = 0$ (type Ia) and $\kappa_\mu \neq 0$ (type Ib). Consider the latter case, i.e. $\kappa_\mu \neq 0$. By virtue of Lemma 1 we may assume that the four matrices κ_μ are in lower triangular form. Then there will be a set $\mathrm{I}_{\mathrm{phys}}$ of values of the index α, defined by the requirement

$$(\kappa_\mu)_{\alpha\beta} = 0 \qquad \text{for} \qquad \alpha \in I_{\mathrm{phys}} \text{ and all } \beta \text{ and } \mu.$$

Consider those components $\varphi_\alpha(x)$ of the fields with $\alpha \in I_{\mathrm{phys}}$. Anticipating later applications we shall refer to these as "physical components". Lemma 1 asserts that such components always exist. Furthermore, from Lemma 1 and proof of Lemma 3, part 1, we know that

$$(\Sigma_{\mu\nu})_{\alpha\beta} = 0, \quad \Delta_{\alpha\beta} = 0 \qquad \text{for} \quad \alpha \in I_{\mathrm{phys}}, \quad \beta \notin I_{\mathrm{phys}}.$$

By Schur's Lemma we may assume, without loss of generality, that

$$\Delta_{\alpha\beta} = -i\lambda_0\delta_{\alpha\beta}; \qquad \alpha, \beta \in I_{\mathrm{phys}}$$

(otherwise there is an invariant subspace on which this is true). Upon inserting these relations into the general transformation law (II.10) it is found that the "physical components" of the field transform under infinitesimal transformations according to the law (III.1)—i.e. a transformation of type Ia—in the sense that e.g. for special conformal transformations

$$\delta\varphi_\alpha(x) = -i\epsilon^\mu \sum_{\beta\in I_{\text{phys}}} \{i(-2lx_\mu + 2x_\mu x_\nu\,\partial^\nu - x^2\,\partial_\mu)\,\delta_{\alpha\beta} - 2ix^\nu(\Sigma_{\mu\nu})_{\alpha\beta}\}\,\varphi_\beta(x);$$

$$\text{for} \quad \alpha \in I_{\text{phys}}\,.$$

An important point here is, of course, that the physical components of the transformed field depend only on the physical components of the old field.

We may summarize our findings in the following

LEMMA 4. *Suppose we are given a representation of the conformal algebra of the form* (II.10), *induced by any finitedimensional representation* κ_μ, $\Sigma_{\mu\nu}$, Δ *of the algebra of the little group* (II.6). *Choose a basis in indexspace as in Lemma 1. Then a suitable (nonempty) part* $\{\varphi_\alpha(x),\ \alpha \in I_{\text{phys}}\}$ *of the components of the field may be selected such that a representation of the form* (III.1) (*i.e. of type* Ia) *is induced on these components.* I_{phys} *is defined by the requirement that*

$$(\kappa_\mu)_{\alpha\beta} = 0 \qquad \textit{for} \quad \alpha \in I_{\text{phys}} \textit{ and all } \beta \textit{ and } \mu. \tag{IV.16}$$

3. *Relation between manifestly covariant fields and fields over Minkowski space.*

After the preparations of the preceding subsection we are finally able to prove the main result of this section.

THEOREM. *Suppose we are given a manifestly conformal covariant (multispinor) field* $\chi(\eta)$, *defined on the cone* $\eta^2 = 0$ *in* R^6, *and satisfying a homogeneity condition*

$$\eta^B\,\partial_B\chi(\eta) = n\chi(\eta) \qquad \textit{where} \qquad \partial_B = \frac{\partial}{\partial\eta^B} \tag{IV.17}$$

Suppose it transforms under infinitesimal conformal transformations according to

$$\delta\chi(\eta) = -i\epsilon^{AB}(L_{AB} + s_{AB})\,\chi(\eta), \qquad L_{AB} \equiv i(\eta_A\partial_B - \eta_B\partial_A)$$

where s_{AB} *is some finite dimensional irreducible representation of* $O(2, 4)$ *algebra acting on indexspace* $\mathscr{H}$. *Let* λ_0 *be the highest eigenvalue of is*$_{65}$. *Consider then*

$$\varphi(x) \equiv \kappa^{-n}\mathrm{E}V\chi(\eta) \tag{IV.18}$$

where

$$x_\mu = \eta_\mu/\kappa; \qquad \kappa = \eta_5 + \eta_6; \qquad V = \exp -i(s_{6\mu} + s_{5\mu})$$

and E is the projection operator onto the eigenspace $\mathscr{H}'$ belonging to eigenvalue λ_0 of is_{65}. We assert

(1) $\delta\varphi(x) = \kappa^{-n}\mathrm{EV}\,\delta\chi(\eta)$ *is given by Eq.* (III.1) *and* (II.3)*f with* $l = n + \lambda_0$ *and real.*

(2) *every field $\varphi(x)$ over Minkowskispace with transformation law* (III.1) *may be obtained in this way from a suitable manifestly covariant field $\chi(\eta)$ as described above, provided the matrices $\Sigma_{\mu\nu}$ in* (III.1) *form an irreducible representation of SL(2C). A suitable SU(2, 2) representation s_{AB} is given in terms of the completely symmetric multispinors described in Lemma* 3.

The extension to reducible $SL(2C)$ representations is obvious. The most important case are the familiar Bargmann Wigner Lorentz-multispinors $\xi_{\{\gamma_1\cdots\gamma_{2j}\}}$, $\gamma_i = 1 \cdots 4$ which transform like a product of $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representations. The conformal extension is given by $\xi_{\{\gamma_1\cdots\gamma_{2j}\}}$ where γ_i now run from $1 \cdots 8$. It transforms under $SU(2, 2)$ like a product of eight-dimensional representations (IV.7).

Proof. The proof of the first part of the theorem will proceed in three steps:

(1) coordinate transformation

(2) x-dependent basis transformation in index space to transform away the intrinsic part of the translation operator, i.e. ensure Eq. (II.8)

(3) Project out unphysical components.

Step 1 has first been carried out by Dirac (*39*). For spin $\frac{1}{2}$ case, step 2 has been described by Hepner (*30*). The present discussion covers arbitrary spin.

Step 1. We introduce a new set of six real independent variables x_μ, κ, η^2 related to η_A by

$$x_\mu = \eta_\mu/\kappa, \qquad \kappa = \eta_5 + \eta_6, \qquad \eta^2 = \eta_A\eta^A$$

Let us now reexpress the functions $\chi_\alpha(\eta)$ as functions of the new variables

$$\chi_\alpha(\eta) \equiv \chi'_\alpha(x_\mu, \kappa, \eta^2)$$

Making use of the chain rule

$$\frac{\partial}{\partial\eta^A} = \left(\frac{\partial x^\mu}{\partial\eta^A}\right)\frac{\partial}{\partial x_\mu} + \left(\frac{\partial\kappa}{\partial\eta^A}\right)\frac{\partial}{\partial\kappa} + \left(\frac{\partial\eta^2}{\partial\eta^A}\right)\frac{\partial}{\partial\eta^2}$$

$$= \frac{1}{\kappa}\left(g_A{}^\mu - [g_A{}^6 - g_A{}^5]\,x^\mu\right)\frac{\partial}{\partial x^\mu} + (g_A{}^6 - g_A{}^5)\frac{\partial}{\partial\kappa} + 2\eta_A\frac{\partial}{\partial\eta^2}$$

we may reexpress $L_{AB} \equiv i(\eta_A \partial_B - \eta_B \partial_A)$ as a differential operator in the new variables. The differential operator $\partial/\partial\eta^2$ does not enter this expression for L_{AB}. L_{AB} is therefore a well defined differential operator acting on functions defined on the cone $\eta^2 = 0$ only. We may then restrict our attention to this cone. We find

$$\eta^A \partial_A = \kappa \frac{\partial}{\partial \kappa} \qquad \text{on the cone } \eta^2 = 0.$$

Hence the homogeneity condition Eq. (IV.4) on $\chi(\eta)$ reads now

$$\left(\kappa \frac{\partial}{\partial \kappa} - n\right) \chi'(x_\mu, \kappa, 0) = 0 \tag{IV.20}$$

Inserting this into the expression for L_{AB} we find finally the action of L_{AB} on the function χ':

$$\begin{aligned} L_{\mu\nu} &= i\left(x_\mu \frac{\partial}{\partial x^\nu} - x_\nu \frac{\partial}{\partial x^\mu}\right) \\ L_{6\mu} + L_{5\mu} &= i \frac{\partial}{\partial x^\mu} \\ L_{6\mu} - L_{5\mu} &= i\left(-2nx_\mu + 2x_\mu x_\nu \frac{\partial}{\partial x_\nu} - x^2 \frac{\partial}{\partial x^\mu}\right) \\ L_{65} &= i\left(-n + x^\nu \frac{\partial}{\partial x^\nu}\right) \end{aligned} \tag{IV.21}$$

Let us now introduce

$$\tilde{\varphi}_\alpha(x) = \kappa^{-n} \chi'_\alpha(x, \kappa, 0) \equiv (\eta_5 + \eta_6)^{-n} \chi_\alpha(\eta) \tag{IV.22}$$

This new field does not depend on κ because of Eq. (IV.20), and is therefore a function of x only as stated. Furthermore a conformal transformation (IV.2) of $\chi(\eta)$ induces on $\tilde{\varphi}(x)$ a transformation

$$\delta\tilde{\varphi}(x) = -i\epsilon^{AB}(L_{AB} + s_{AB})\, \tilde{\varphi}(x) \tag{IV.23}$$

where L_{AB} is given by Eq. (IV.21). To see this, note that L_{AB} does no longer contain any differentiation $\partial/\partial\kappa$ and the proportionality factor κ^{-n} in Eq. (IV.22) therefore commutes with L_{AB}. In particular we have from (IV.21) and (IV.23)

$$P_\mu \tilde{\varphi}(x) = \left(i \frac{\partial}{\partial x^\mu} + \pi_\mu\right) \qquad \text{where} \qquad \pi_\mu = s_{6\mu} + s_{5\mu}$$

Step 2. Define (*11, 30*)

$$\varphi_\alpha(x) = V_{\alpha\beta} \tilde{\varphi}_\beta(x) \qquad \text{where} \quad V = \exp -ix^\nu \pi_\nu \tag{IV.24}$$

V is always a finite polynomial in x because the π_μ are also nilpotent. It could therefore be worked out explicitly in each case. In practice such straightforward but sometimes tedious calculations can usually be avoided by using translation invariance and the fact that $V = 1$ at $x = 0$.

Because all π_μ commute, V has an inverse given by

$$V^{-1} = \exp(+ix^\nu \pi_\nu)$$

Furthermore

$$V\left(i\frac{\partial}{\partial x^\mu}\right)V^{-1} = i\frac{\partial}{\partial x^\mu} - \pi_\mu$$

Using Eqs. (IV.21) and the C.R. of the matrices s_{AB} as given by (II.4) one may check that the components of the field φ in the new basis do indeed transform according to Eq. (II.10) with

$$\Sigma_{\mu\nu} = s_{\mu\nu}\,; \qquad \kappa_\mu = s_{6\mu} - s_{5\mu}\,; \qquad \Delta = s_{65} - in$$

n is given by Eq. (IV.4). The remaining matrices have disappeared from the transformation law.

Step 3. After having arrived at a field which transforms according to Eq. (II.10) it remains to project out unphysical components such that the term involving κ_μ in Eq. (II.10) goes away. This problem has been solved by Lemma 4 in the preceding subsection.

This completes the proof of the first part of the theorem. The second part is now a consequence of the first, and of Lemmas 3 and 4.

4. *Invariant wave equations and interactions.*

With manifestly covariant fields it is straightforward to write down manifestly invariant wave equations and interactions. The following examples are due to Kastrup (*40*).

All fields are to correspond to fields over Minkowski space which transform according to Eq. (II.10) with $\kappa_\mu = 0$, as do the fields employed in Section III.

The spin 0 field (scalar or pseudoscalar) corresponds to a conformal scalar $A(\eta)$ with degree of homogeneity $n = -1$, i.e.

$$\eta^B \partial_B A(\eta) = -A(\eta), \qquad s_{AB}A(\eta) \equiv 0.$$

The free wave equation is

$$\Box_6 A(\eta) = 0 \qquad \text{where} \quad \Box_6 = \partial^B \partial_B\,. \tag{IV.25}$$

As discussed by Dirac (*39*), this is a well defined equation for $A(\eta)$ defined on the cone $\eta^2 = 0$ only, if and only if $n = -1$ as we assume. By Eq. (IV.18)f, $l = n = -1$ in agreement with the discussion in Section III. The scalar field in Minkowski space is then given by

$$a(x) = (\eta_5 + \eta_6)\, A(\eta)$$

and satisfies

$$\Box_x a(x) = 0.$$

The spin $\frac{1}{2}$ field is an 8-component spinor $\chi(\eta)$ of degree of homogeneity $n = -2$

$$\eta^B \partial_B \chi(\eta) = -2\chi(\eta), \qquad s_{AB} = \gamma_{AB} \text{ given by Eq. (IV.7).} \tag{IV.26}$$

The adjoint is defined by

$$\bar{\chi} = \chi^* \gamma_0 \tau_1 \tag{IV.27}$$

cf. Eq. (IV.10).

The corresponding 8-spinor over Minkowski space is again given by Eqs. (IV.18) which reads

$$\psi(x) = (\eta_5 + \eta_6)^{+2}\,(1 + ix^\mu \gamma_\mu \tau^-)\,\chi(\eta) \quad \text{where} \quad \tau^- = \tfrac{1}{2}(\tau_1 - i\tau_2) \tag{IV.28}$$

Its physical components are those satisfying

$$(1 + \tau_3)\,\psi(x) = 0. \tag{IV.29}$$

In the basis where τ_3 has the usual diagonal form, these are just the lowest four components. From Eqs. (IV.7) and (IV.18)f one finds $l = n + \frac{1}{2} = -\frac{3}{2}$, as was assumed in Section III because of unitarity requirements.

The free wave equation is

$$-2i(\gamma^{AB} L_{AB} + 2)\,\chi(\eta) = 0; \qquad L_{AB} \equiv i(\eta_A \partial_B - \eta_B \partial_A). \tag{IV.30}$$

This amounts to diagonalizing the second-order Casimir operator $\frac{1}{2} J^{AB} J_{AB}$ ((Hepner and Murai) (*30*, *42*)).

The spin 1 gauge fields are 6-vectors $A_B(\eta)$ of degree of homogeneity $n = -1$.

$$\eta^B \partial_B A_C(\eta) = -A_C(\eta); \qquad (s_{AB} A)_C = i(g_{AC} A_B - g_{BC} A_A) \tag{IV.31}$$

satisfying the subsidiary condition

$$\eta^B A_B(\eta) = 0. \tag{IV.32}$$

If we impose in addition the generalized Lorentz condition

$$\partial^C A_C(\eta) = 0$$

then the admissible gauge transformations are, for the electromagnetic potential,

$$A_C(\eta) \to A_C(\eta) + \partial_C S(\eta) \tag{IV.33}$$

where the gauge function S must be specified on a whole neighbourhood of the cone $\eta^2 = 0$, and satisfy there

$$\eta^B \partial_B S(\eta) = 0; \qquad \Box_6 S(\eta) = 0.$$

The free field equation is then

$$\Box_6 A_C(\eta) = 0. \tag{IV.34}$$

Again, the choice of $n = -1$ makes this into a well-defined equation for $A_C(\eta)$ defined on the cone $\eta^2 = 0$ only.

The corresponding field $a_B(x)$ is again given by Eqs. (IV.18)

For the first four components this takes the explicit form

$$a_\mu(x) = (\eta_5 + \eta_6)[A_\mu(\eta) - x_\mu\{A_5(\eta) - A_6(\eta)\}] \qquad \mu = 0 \cdots 3 \tag{IV.35}$$

and the subsidiary condition Eq. (IV.32) reads

$$a_6(x) - a_5(x) = 0 \tag{IV.36}$$

This can be seen in the following way: For $x_\mu = 0$, we have $\eta_\mu = 0$ and $\eta_5 = \eta_6$. Therefore Eq. (IV.36) is clearly true at this point. Now Eq. (IV.32) is conformal invariant and therefore, in particular, translation invariant. However, by construction, all $a_B(x)$ transform under translations according to Eq. (II.10). Thus, by translation invariance the validity of Eq. (IV.36) for arbitrary x_μ follows from its validity at $x_\mu = 0$.

The first four components of $a_B(x)$ are the physical ones as they satisfy Eq. (IV.16) by virtue of the subsidiary condition Eq. (IV.36):

$$(\kappa_\mu a)_\nu(x) = (\{s_{6\mu} - s_{5\mu}\}a)_\nu = -ig_{\mu\nu}(a_6 - a_5) = 0 \qquad \nu = 0 \cdots 3 \tag{IV.37}$$

Invariant couplings. Following Kastrup, it is easy to see that a conformal invariant coupling between a pseudoscalar field $A(\eta)$ and the spin $\frac{1}{2}$ field $\chi(\eta)$ is given by the following wave equation:

$$\begin{aligned} \Box_6 A &= -g\eta^C \bar{\chi}\beta_C\beta_7\chi \\ -2i(\gamma^{AB}L_{AB} + 2)\chi &= g\eta^C\beta_C\beta_7\chi A \end{aligned} \tag{IV.38}$$

and the coupling of the electromagnetic field to the spin $\frac{1}{2}$ field is given by

$$[2\gamma^{AB}\{\eta_A(\partial_B - iqA_B) - \eta_B(\partial_A - iqA_A)\} - 4i]\chi = 0$$
$$\Box_6 A_C = qj_C(\eta) \qquad \text{(IV.39)}$$

where

$$j_C(\eta) = 2\eta^B\bar{\chi}\gamma_{BC}\chi.$$

The β-matrices are given by Eqs. (IV.8)f.

As is seen, the electromagnetic coupling is obtained by making the gauge-invariant substitution $\partial_C \to \partial_C - iqA_C$. In this form it can be immediately generalized to arbitrary sets of gauge fields $A_C{}^a(\eta)$. Let $\mathbf{T}_a$ be the representation matrices of the relevant group as discussed in Section III; then the general rule is to substitute

$$\partial_C \to \partial_C - iqA_C{}^a\mathbf{T}_a \qquad \text{(IV.40)}$$

in the free field equations. Summation over a is understood. In this way one obtains couplings which are both conformal invariant and gauge invariant.

Finally there also exists a quadrilinear conformal invariant spin 0 boson coupling. A corresponding wave equation would be, e.g.,

$$\Box_6 A(\eta) = g[A(\eta)]^3. \qquad \text{(IV.41)}$$

Of course all couplings mentioned above can occur simultaneously.

A point which we wish to stress is the following: Not all manifestly covariant looking couplings are actually acceptable. They must in addition satisfy the following requirements:

(1) The wave equations must have homogeneous functions as solutions. This requires that all terms in a certain wave equation must have the same degree of homogeneity. The degree of homogeneity of such a term is calculated by counting each explicit coordinate η with $+1$, each derivative $\partial/\partial\eta$ with -1, and each field with the appropriate number n indicating its degree of homogeneity (e.g. $n = -1$ for scalar and vector, and -2 for 8-spinors χ in the cases discussed above.)

(2) The interaction terms must not couple unphysical field components to physical ones.

The couplings given above satisfy this condition, while e.g. a coupling $\bar{\chi}\chi A$ would violate (2).

Condition (2) is crucial and may not be given up without destroying the physical interpretation of the theory. It *must* be satisfied at least for a suitable choice of the electromagnetic gauge (IV.33). A sufficient condition for this is invariance of the wave equation under a new type of additive gauge transformations which affect

only the unphysical components. For the spin $\frac{1}{2}$ field such gauge transformations have the form

$$\chi'(\eta) = \chi(\eta) + \eta^A \beta_A \zeta(\eta), \qquad \eta^A \partial_A \zeta = (n-1)\zeta,$$

where ζ is an arbitrary complex 8-spinor function like χ, and satisfies the indicated homogeneity condition. β_A is given by Eq. (IV.8) or (IV.9). Invariance of Eqs. (IV.38), (IV.39) follows immediately from $(\eta^A \beta_A)^2 = 0$, and Eqs. (IV.10) and (IV.32).

There is also an easy way of checking by explicit computation whether condition (2) is satisfied, without going through the tedious transformations of Section (IV.3). Because of translation invariance, it is sufficient to check that the condition is satisfied at $x_\mu = 0$. This corresponds to $\eta_\mu = 0$ and $\eta_6 = \eta_5$. At this point the boost operator V in Eq. (IV.18) is simply unity: $V = 1$.

By using this trick it is also easy to verify that the wave equations (IV.38) and (IV.39) do indeed correspond to the Dirac, Klein-Gordon and Maxwell equations for the physical field components in Minkowski space, with minimal electromagnetic interaction and nonderivative pseudoscalar pion-nucleon interaction.

In Section III an alternative characterization of all (physically acceptable) conformal invariant couplings for spin $\leqslant 1$ has been given. If all vector mesons are assumed to be gauge fields, then the manifestly conformal invariant couplings given above, and their obvious generalization to the case of several fields of the same spin, exhaust all possibilities for spin $\leqslant 1$. This may be checked by enumerating all possibilities, as there are only a few types of couplings with dimensionless coupling constant, and the gauge field couplings are fixed in their form.

V. REPRESENTATIONS INDUCED BY REPRESENTATIONS OF THE LITTLE GROUP WITH $\kappa_\mu \neq 0$.

In this Section we review an attempt to understand $V - A$ or $V + A$ weak interaction theory as a conformally invariant theory of fundamental interactions. This is at the same time as yet the only application which representations with $\kappa_\mu \neq 0$ have found. To the best of our knowledge, the idea goes back to Hepner (*30*). In the sequel we follow Hepner in part.

Assume we are working with a four component spinor (quark, μ or e-field). If we postulate that a four fermion interaction be invariant under conformal transformations (II.10) with (IV.6), then there exist two possible couplings

$$g\bar{\psi}_1\psi_2\bar{\psi}_3\psi_4 \tag{V.1a}$$

$$g\bar{\psi}_1\gamma_\mu(1 \pm i\gamma_5)\,\psi_2\bar{\psi}_3\gamma^\mu(1 \pm i\gamma_5)\,\psi_4 \tag{V.1b}$$

where we have $(1 \pm i\gamma_5)$ depending on whether $\kappa_\mu = \mp \frac{1}{2}\gamma_\mu(1 \pm i\gamma_5)$, i.e. which of

the two four-dimensional representations $\Delta^{(\pm)}$ we use for the little group. (These couplings correspond to $\bar{\chi}_1\chi_2\bar{\chi}_3\chi_4$ and $\eta_A\bar{\chi}_1\gamma^{AB}\chi_2\eta^C\bar{\chi}_3\gamma_{CB}\chi_4$ in the six dimensional language of Section IV.) The most general conformal invariant linear field equation with interaction (V.1b) is determined to be

$$\left\{i\gamma_\mu \tfrac{1}{2}(1 \pm i\gamma_5)\,\partial^\mu + \frac{i}{2}(n-2)\,\gamma_5 - m\right\}\psi = g\gamma_\mu(1 \pm i\gamma_5)\,\psi\bar{\psi}\gamma^\mu(1 \pm i\gamma_5)\psi \qquad \text{(V.2)}$$

Here $\Delta = n \mp \frac{1}{2}\gamma_5$, this relates n to Δ which appears in the transformation law (II.10). It turns out that it must satisfy $n^3 - n + 2 = 0$ in order that Eq. (V.2) be actually invariant; therefore $n \neq 2$.

While the interaction term is precisely what we want, the LHS of Eq. (V.2) is not. For one thing the term $\frac{1}{2}i(n-2)\gamma_5 \neq 0$ prevents us from having a conserved hermitian probability current. Secondly, those components of ψ which are not coupled do not contribute to the kinetic energy in Eq. (V.2) either. Therefore, if we want to keep all four components of our spinor field as physical we have no choice but to add suitable symmetry breaking terms to the kinetic energy in Eq. (V.2) such that we arrive at the usual Dirac equation with weak interaction:

$$(i\gamma_\mu\partial^\mu - m)\psi = g\gamma_\mu(1 \pm i\gamma_5)\,\psi\bar{\psi}\gamma^\mu(1 \pm i\gamma_5)\psi \qquad \text{(V.3)}$$

In other words, the interaction term in Eq. (V.3) is conformal invariant, but the kinetic energy term is not. Of course, we thereby loose the existence of hermitian dilatation and conformal currents, as these rely on the invariance of the canonical equal time C.R. of the fields with their conjugate momenta, and hence of the kinetic energy (apart from mass). This is in contrast with the theory of Section III which is different from the one presently discussed both in physical spirit and mathematical detail.

APPENDIX: THE σ-MODEL AS AN ILLUSTRATION OF IDEAS IN SECTION III

Consider the Lagrangian of the σ-model of Gell-Mann and Lévy (*43*)

$$\mathscr{L} = \bar{N}(i\not{\partial} - M)\,N + ig\bar{N}\gamma_5\tau N\pi + \tfrac{1}{2}(\partial_\mu\pi\,\partial^\mu\pi - \mu^2\pi^2)$$
$$+ \frac{1}{2}\left(\partial_\mu\sigma\partial^\mu\sigma - \left\{\mu^2 + \frac{2\lambda}{f^2}\right\}\sigma^2\right) - \lambda\left(\{\pi^2 + \sigma^2\} - \frac{2}{f}\sigma\{\pi^2 + \sigma^2\}\right). \qquad \text{(A.I)}$$

Here N is the nucleon field, π is the pion field, and σ is the field of a $I = 0$, $J^P = 0^+$ meson. $f = g/2M$.

Let us choose the free parameter λ to be

$$\lambda = g^2 \frac{\mu^2}{4M^2}$$

Reexpressing the Lagrangian in terms of the field $\sigma' = \sigma - (2f)^{-1}$, and calculating the dilatation current and conformal currents from Eq. (III.3), one finds for their divergences

$$\text{(a) } \partial^\mu \mathscr{D}_\mu = g^{-1} M m_\sigma{}^2 \sigma(x), \qquad \text{(b) } \partial^\nu{}_\nu \mathscr{K}_\mu = 2x_\mu \partial^\nu \mathscr{D}_\nu . \tag{A.2}$$

The last equation follows direct from Eq. (III.10)ff since the present Lagrangian does not involve any derivative couplings. m_σ is the (bare) σ-mass. *We see that in the limit of a massless boson σ both currents are conserved, and we have a spontaneous breakdown of conformal symmetry.*

With the usual definition of the axial vector current $\mathscr{A}_j{}^\mu$ for this model, one finds that generally, also for $m_\sigma \neq 0$

$$[F_i{}^5(x_0), \partial_\mu \mathscr{A}_j{}^\mu(x)] = -i\delta_{ij} \tfrac{1}{2}(\partial^\mu \mathscr{D}_\mu - \tfrac{3}{4}\mu^2 f^{-2}) \qquad \text{for} \quad i, j = 1, 2, 3$$

where

$$F_i{}^5 = \int d^3x \mathscr{A}_i{}^0. \tag{A.3}$$

Elsewhere it has been proposed to generalize this formula to chiral $SU(3) \otimes SU(3)$ in the following form (*16*):

$$\partial^\mu \mathscr{D}_\mu = \alpha_0 u_0(x) + \alpha_8 u_8(x) - \langle 0 \mid \{\alpha_0 u_0 + \alpha_8 u_8\} \mid 0 \rangle \tag{A.4}$$

with

$$\alpha_0 + \frac{1}{\sqrt{2}} \alpha_8 = -3\sqrt{\frac{3}{2}}$$

α_8/α_0 is a measure of the breaking of the eightfold way. The u_i must satisfy the C.R. of (integrated) scalar densities with vector and axial vector currents as proposed by Gell-Mann ($i, j, k = 0 \cdots 8$) (*44*)

$$\begin{aligned} [F_i{}^5, v_j(x)]_{\text{eq.t.}} &= i\, d_{ijk} u_k(x) \\ [F_i{}^5, u_j(x)]_{\text{eq.t.}} &= -i\, d_{ijk} v_k(x) \end{aligned} \tag{A.5}$$

with

$$v_j(x) = \partial^\mu \mathscr{A}_\mu{}^j(x) \qquad \text{for} \quad j = 1, 2, 3.$$

The matrix elements of $u_0 + 1/\sqrt{2}\, u_8$ are known in current algebra calculations as "σ-terms." A method to calculate them on the basis of Eq. (A.4) and Eq. (III.1) has been outlined in Ref. (*16*).

Finally we wish to point out that the "σ-meson" plays two different roles in this model. Firstly it is a manifestation of the breaking of dilatation symmetry, Eq. (A.2a). Secondly, it provides an attractive πN force which is necessary to cancel most of the big s-wave repulsion inherent in the nonderivative πN coupling, which is not observed experimentally. Recall that requirement of the validity of Eq. (A.2b) does not allow for a derivative πN coupling (Section III). There seems to be experimental evidence for the existence of a scalar, isoscalar meson with a mass around 700 MeV (*45*).

RECEIVED: January 2, 1969

REFERENCES

1. F. GÜRSEY, *Nuovo Cimento* **3**, 988 (1956).
2. H. A. KASTRUP, *Ann. Physik* **7**, 388 (1962).
3. T. FULTON, R. ROHRLICH AND L. WITTEN, *Rev. Mod. Phys.* **34**, 442 (1962).
4. Y. MURAI, *Progr. Theoret. Phys.* (Kyoto) **9**, 147 (1953).
5. A. ESTEVE AND P. G. SONA, *Nuovo Cimento* **32**, 473 (1964).
6. I. M. GEL'FAND AND M. I. GRAEV, *Izv. Akad. Nauk SSSR, Ser. Mat.* **29**, 1329 (1965) and "Proceedings of the international spring school for theoretical physics," Yalta (1966).
7. A. KIHLBERG, V. F. MULLER, AND F. HALBWACHS, Commun. Math. Phys. **3**, 194 (1966).
8. I. T. TODOROV, *ICTP, Trieste*, preprint IC/66/71.
9. R. RAĆZKA, N. LIMIÇ, AND J. NIEDERLE, *J. Math. Phys.* **7**, 1861, 2026 (1966); **8**, 1079 (1967).
10. T. YAO, *J. Math. Phys.* **8**, 1931 (1967); **9**, 1615 (1968).
11. R. L. INGRAHAM, *Nuovo Cimento* **12**, 825 (1954).
12. L. GROSS, *J. Math. Phys.* **5**, 687 (1964).
13. L. CASTELL, *Nucl. Phys.* **B4**, 343 (1967).
14. B. KURŞUNOGLU, *J. Math. Phys.* **8**, 1694 (1967).
15. G. MACK AND I. TODOROV, *ICTP, Trieste*, preprint IC/68/86.
16. G. MACK, *Nucl. Phys.* **B5**, 499 (1968).
17. G. MACK, *Phys. Letters* **26B**, 515 (1968).
18. Ref. (*15*), Appendix.
19. A. GAMBA AND G. LUZATTO, *Nuovo Cimento* **18**, 1086 (1960).
20. E. C. ZEEMAN, *J. Math. Phys.* **5**, 490 (1964).
21. T. FULTON, R. ROHRLICH, AND L. WITTEN, *Nuovo Cimento* **26**, 652 (1962) and Ref. (*3*).
22. L. CASTELL, Nucl. Phys. **B5**, 601 (1968); *Nuovo Cimento* **46A**, 1 (1966).
23. J. ROSEN, Brown University preprints NYO-2262 TA-151, NYO-2262 TA-161 (1967).
24. H. A. KASTRUP, *Phys. Rev.* **142**, 1060 (1966); **143**, 1041 (1966); **150**, 1189 (1966); *Nucl. Phys.* **58**, 561 (1964) and Ref. (*2*).
25. M. FLATO AND D. STERNHEIMER, *Compt. Rend.* **263**, 935 (1966).
26. D. BOHM, M. FLATO, D. STERNHEIMER, AND J. P. VIGIER, *Nuovo Cimento* **38**, 1941 (1965).
27. J. E. WESS, *Nuovo Cimento* **18**, 1086 (1960).

28. G. W. MACKEY, *Bull. Am. Math. Soc.* **69**, 628 (1963); R. HERMANN, "Lie groups for physicists", (W. A. Benjamin), New York (1966), ch. 9.
29. L. O'RAIFERTAIGH, *Phys. Rev. Letters* **14**, 575 (1965).
30. N. JACOBSON, "Lie algebras," Interscience, New York (1962), p. 45, Corollary 2.
31. W. A. HEPNER, *Nuovo Cimento* **26**, 352 (1962).
32. G. MACK, Ph.D. thesis, Berne, Switzerland (1967).
33. J. A. MCLENNAN, *Nuovo Cimento* **3**, 1360 (1956); **5**, 640 (1957).
34. C. N. YANG AND R. L. MILLS, *Phys. Rev.* **96**, 531 (1954).
35. S. WEINBERG, *Phys. Rev.* **134**, B882 (1964); **138**, B988 (1965).
36. ABDUS SALAM AND J. C. WARD, *Phys. Letters* **13**, 168 (1964).
37. D. M. GREENBERGER, *Ann. Phys. (N.Y.)* **25**, 290 (1963).
38. R. UTIYAMA, *Phys. Rev.* **101**, 1597 (1956).
39. P. A. M. DIRAC, *Ann. Math.* **37**, 429 (1936).
40. H. A. KASTRUP, *Phys. Rev.* **150**, 1186 (1966).
41. M. HAMERMESH, "Group theory", Addison–Wesley, London (1962).
42. Y. MURAI, *Nucl. Phys.* **6**, 489 (1958).
43. M. GELL-MANN AND M. LEVY, *Nuovo Cimento* **16**, 705 (1960).
44. M. GELL-MANN, *Phys. Rev.* **125**, 1067 (1962).
45. S. MARATECK *et al.*, *Phys. Rev. Letters* **21**, 1631 (1968).

Reprinted from THE PHYSICAL REVIEW, Vol. 184, No. 5, 1750–1759, 25 August 1969

Nonlinear Realizations. I. The Role of Goldstone Bosons

ABDUS SALAM* AND J. STRATHDEE
International Atomic Energy Agency, International Centre for Theoretical Physics, Miramare, Trieste, Italy
(Received 26 December 1968)

A method, based upon analogies with the Wigner boost technique, is presented for setting up the nonlinear realizations of any continuous symmetry group. It is argued that such realizations are relevant to and useful in the treatment of cases of spontaneous symmetry breakdown.

1. INTRODUCTION

THE equations of physics are usually expressed in a form showing manifest covariance under the transformations of the space-time and internal-symmetry groups. Recently, however, some attention has been paid to the possibility of expressing internal symmetries of the chiral type in a fashion which is not manifestly covariant, some of the transformations of these groups being realized nonlinearly.[1] In the present paper the nonlinear method is shown to provide an economical framework within which to treat the problems of spontaneously broken symmetries,[2] i.e., symmetries which are present in the Lagrangian but not in the ground state. Beyond this, the view will be presented that the nonlinear method, as a method embodying dynamics rather than pure group theory, is applicable *only* to situations in which the symmetry is broken spontaneously. Much of the following discussion will be in the nature of a review. However, one purpose of this is to present the notational developments that emerge through a consistent use of the language of Wigner boosts which, it seems to us, greatly clarifies the subject of nonlinear realizations. We shall need this development in particular in Paper II, where we consider nonlinear realizations of the conformal group in space-time.

The discussions in Secs. 1 and 2 concern the general properties of nonlinear realizations. These have been dealt with in great generality elsewhere.[3] However, in order to make clear the notation adopted in this paper and in Paper II, it will be advantageous to go over these matters in some detail. Moreover, this review will serve to introduce our main idea, which is that the nonlinear realizations are intimately involved in situations where the symmetry breaking is spontaneous. This idea will be further developed in Sec. 3. Finally, to illustrate the method, Sec. 4 is devoted to the nonlinear realizations of $SU(3)$, which become linear with respect to the subgroup $SU(2)\times U(1)$. Thus, the material of Secs. 1 and 2 consists of known results expressed in a new notation, while that of Secs. 3 and 4 is essentially new.

Central to the method of nonlinear realizations is the notion of a *preferred field* that provides a bridge to representations which are linear but *constrained*. There is in fact a complete duality between sets of fields, on the one hand, which transform linearly while being subject to certain nonlinear constraints and, on the other hand, equivalent sets of unconstrained fields which transform according to nonlinear realizations. The preferred field is used in the formation of covariant equations of constraint upon the set of fields which transforms linearly. If these constraints are then used to eliminate all dependent components, there results what is commonly called a nonlinear realization. More specifically, the realization will be linear with respect to all but the preferred field itself, which generally enters nonlinearly. The structure of the realizations arrived at in this way must, of course, depend critically upon the particular set of fields which are chosen to play the preferred role. This choice depends in turn upon the nature of the vacuum symmetry since, as will be seen in the following, the preferred fields are neither more nor less than the field of the Goldstone bosons.[4]

A rather trivial and very familiar example of the procedure just outlined is provided in the case of chiral $SU(2)\times SU(2)$.[4] Let the chiral four-vector π_α play the role of preferred field and let it be employed in the formulation of a set of algebraic constraints upon the fields ϕ_α and $F_{\alpha\beta}$, a chiral four-vector and six-tensor, respectively. For the equations of constraint one might take, for example,

$$\begin{aligned}&\pi_\alpha\phi_\beta-\pi_\beta\phi_\alpha=fF_{\alpha\beta},\\&\pi_\alpha F_{\alpha\beta}-f\phi_\beta=0,\end{aligned}\tag{1.1}$$

where f denotes a numerical constant. The question of how such constraints could arise within the context of a dynamical model is not, for the present, at issue.

* On leave of absence from Imperial College, London, England.

[1] For a recent compilation of the very numerous references to nonlinear realizations see the review by S. Weinberg, in *Proceedings of the Fourteenth International Conference on High-Energy Physics, Vienna, 1968* (CERN, Geneva, 1968), p. 253.

[2] J. Goldstone, Nuovo Cimento 19, 154 (1961); J. Goldstone, Abdus Salam, and S. Weinberg, Phys. Rev. 127, 965 (1962); T. W. B. Kibble, in *Proceedings of the Conference on Particles and Fields, Rochester, 1967* (Wiley-Interscience, Inc., New York, 1967).

[3] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. 177, 2239 (1969); C. G. Callan, S. Coleman, J. Wess, and B. Zumino, *ibid.* 177, 2247 (1969); C. J. Isham, Imperial College, London, Report No. ICTP/67/25 (unpublished).

[4] We wish to emphasize that this paper is *not* about chiral symmetries, which are discussed in this section as an illustration only. Our interest is with the general method of nonlinear realizations, of which the chiral realizations are one particular example.

It follows from (1.1) that, in particular,

$$\pi_\alpha\pi_\alpha=f^2,\qquad \pi_\alpha\phi_\alpha=0, \tag{1.2}$$

so that the 16 various components can be expressed in terms of six independent ones, say, π and ϕ. The linear transformation laws appropriate to the chiral four-vectors π_α and ϕ_α are exemplified by

$$\begin{aligned}&\pi_{1,2}\to\pi_{1,2},\\ &\pi_3\to\pi_3\cos\omega-\pi_4\sin\omega,\\ &\pi_4\to\pi_3\sin\omega+\pi_4\cos\omega,\end{aligned} \tag{1.3}$$

corresponding to a purely chiral transformation. It remains only to eliminate the dependent components π_4 and ϕ_4 by means of the constraints (1.2). The independent components π and ϕ then transform according to *nonlinear realizations which are exemplified by*

$$\begin{aligned}&\pi_{1,2}\to\pi_{1,2},\\ &\pi_3\to\pi_3\cos\omega-(f^2-\pi^2)^{1/2}\sin\omega,\end{aligned} \tag{1.4}$$

and

$$\begin{aligned}&\phi_{1,2}\to\phi_{1,2},\\ &\phi_3\to\phi_3\cos\omega+[\phi\cdot\pi/(f^2-\pi^2)^{1/2}]\sin\omega,\end{aligned} \tag{1.5}$$

and one sees that only the preferred field π enters nonlinearly. As is well known, the form of this result depends very much on the manner in which π_α is parametrized. An alternative scheme would present π_α in the form

$$\begin{aligned}&\pi_i=\varphi_i/(1+\varphi^2/f^2)^{1/2},\quad i=1,2,3\\ &\pi_4=f/(1+\varphi^2/f^2)^{1/2},\end{aligned} \tag{1.6}$$

where the three-vector φ_i is independent. Corresponding to the purely chiral transformations (1.3), one finds for φ_i the transformation law

$$\begin{aligned}&\varphi_{1,2}\to\frac{\varphi_{1,2}}{\cos\omega+(\varphi_3/f)\sin\omega},\\ &\varphi_3\to\frac{\varphi_3\cos\omega-f\sin\omega}{\cos\omega+(\varphi_3/f)\sin\omega},\end{aligned} \tag{1.7}$$

and, for ϕ_i,

$$\begin{aligned}&\phi_{1,2}\to\phi_{1,2},\\ &\phi_3\to\phi_3\cos\omega+(1/f)\phi\cdot\varphi\sin\omega,\end{aligned} \tag{1.8}$$

which replace (1.4) and (1.5), respectively. The realization (1.6) was adopted by Schwinger.[5] We have reproduced it here because it lends itself readily to "proving" the noninvariance of the vacuum. Thus, for infinitesimal ω the transformation (1.7) gives, in particular,

$$[Q,\varphi_3]=-\delta\varphi_3=\omega f(1+\varphi_3^2/f^2), \tag{1.9}$$

the vacuum expectation value of which cannot possibly vanish[6] since the right-hand side is positive-definite. The realization (1.6) *is possible only if the vacuum is degenerate.* Notice that this argument does not depend on the existence or nonexistence of a Lagrangian.

For alternative parametrizations such as (1.4), it is not possible to make such a categorical statement. However, for practical purposes where the nonlinearities are always interpreted by power-series expansions in π/f, the implication is the same. Nonlinear constraints can be dealt with by power-series methods only in theories with degenerate vacuums. The particles associated with the preferred fields are, in fact, the Goldstone bosons. One way to see this is to remark that nowhere in a chiral-invariant Lagrangian does the field π appear without being accompanied by $\partial_\mu\pi$ as well. It follows that the fields π must describe massless (spin-zero) particles. Thus a vacuum state is indistinguishable from a state with two zero-frequency pions, four zero-frequency pions, etc., provided the pions together form an $I=0$ multiplet of the subgroup $SU(2)$. Now, since the pions form an incomplete multiplet of $SU(2)\times SU(2)$, it is clear that these physically degenerate and indistinguishable states of lowest energy will lead to a (spontaneous) breakdown of chiral symmetry.

So far we have said that if nonlinear realizations are introduced by considering linear realizations of the preferred fields together with a constraint, the constraint implies that the vacuum state in the theory must be a noninvariant state, and the symmetry a spontaneously broken one, with the independent ones among the preferred fields playing the role of Goldstone bosons. Consider now the converse problem: Given a theory with a chiral-invariant Lagrangian it may happen that the ground state is not chiral-invariant. In this case there must appear Goldstone bosons corresponding to the components of the symmetry which are absent from the vacuum. One may introduce into the theory a set of spin-zero fields describing these bosons. The problem one is presented with is how to couple such mesons with other particles so that the consequences of the vacuum asymmetry are made explicit. These effects must include, for example, the guarantee that these particles remain massless even after interacting with other particles. The formalism must also give a correct account of the perturbations of masses and coupling constants on account of symmetry breaking. *Our solution to the problem is to suggest that the appropriate formalism is the one where a nonlinear realization is employed with these spin-zero fields playing the role of the preferred fields.*

The problem of making explicit the Goldstone bosons can be solved in various ways.[7] Suppose, for simplicity,

[5] J. Schwinger, Phys. Letters **24B**, 473 (1967).

[6] S. Gasiorowicz and D. Geffen in their excellent review of chiral symmetries [Argonne National Laboratory Report No. ANL/HEP 6809 (unpublished)] make this point though they do not stress how vital it is for the dynamical relevance of the whole nonlinear method.

[7] T. W. B. Kibble, Phys. Rev. **155**, 1554 (1967).

that the Lagrangian contains a zero-spin chiral four-vector Φ_α; then it is possible to eliminate Φ_4 in favor of $\chi=(\Phi_\alpha\Phi_\alpha)^{1/2}$. The kinetic energy then takes the form

$$\tfrac{1}{2}(\partial_\mu\Phi_\alpha\partial_\mu\Phi_\alpha-m^2\Phi_\alpha\Phi_\alpha)=\tfrac{1}{2}(\partial_\mu\chi\partial_\mu\chi-m^2\chi^2)+\tfrac{1}{2}D_\mu\Phi\cdot D_\mu\Phi, \quad (1.10)$$

where $D_\mu\Phi$ denotes the so-called covariant derivative of Φ. It is given by

$$\frac{1}{\chi}D_\mu\Phi=\partial_\mu\left(\frac{\Phi}{\chi}\right)-\frac{\Phi}{\chi}\times\partial_\mu\left(\frac{\Phi}{\chi}\right) \quad (1.11)$$

and belongs to the nonlinear realization (1.5). The second term in (1.10) contains the term $\frac{1}{2}(\partial_\mu\Phi)^2$ together with an infinite number of interaction terms which arise from the expansion of $1/\chi$ in powers, i.e.,

$$\frac{1}{\chi}=\frac{1}{\langle\chi\rangle+\chi'}=\frac{1}{\langle\chi\rangle}-\frac{\chi'}{\langle\chi\rangle^2}+\frac{\chi'^2}{\langle\chi\rangle^3}-\cdots,$$

which is meaningful provided $\langle\chi\rangle\neq0$. The nonvanishing of $\langle\chi\rangle$ must be thought of as a consequence of the supposed vacuum asymmetry.[8]

The particle which is characterized by the isoscalar field χ' is of no particular importance in the theory. It arose as a by-product of the effort to set up an effective Lagrangian with Goldstone bosons. Having obtained the effective Lagrangian, one is at liberty to set $\chi'=0$. Only the numerical part $\langle\chi\rangle$ must be kept.

An important exception to the theorem which requires the presence of massless bosons in situations where a symmetry is broken intrinsically occurs when long-range vector fields are also present.[9] If the currents of the spontaneously broken symmetries are coupled to a gauge field of the Yang-Mills type, then the symmetry breaking manifests itself not through the appearance of Goldstone bosons but rather in the acquiring of mass by some of the components of the gauge field. This phenomenon, which was discovered by Anderson and developed by Higgs and Kibble, will be presented in the nonlinear notation in Sec. 2 where it will be shown that the preferred field disappears from the Lagrangian if the symmetry group is gauged. In addition—and this is where we improve on Higgs and Kibble—the residual quantized objects such as χ', whose existence is required in their models, can be set equal to zero without doing any violence to the elegant formulation afforded by the nonlinear formalism.

Turning back to the formulation of the nonlinear method, it will be remarked that, in Eqs. (1.1), the field π_α plays a role which is formally similar to that taken by the four-momentum p_α in the formulation of relativistically covariant free-field equations. The analogy can be extended to interaction terms as well. The involvement of orbital angular momentum in the relativistic coupling of particles with spin is accounted for by the presence of terms such as $\gamma_\alpha\partial/\partial x_\alpha$ in the Lagrangian. Likewise, the involvement of soft pions in the chiral-invariant coupling of particles with isospin is brought about through terms such as $\Gamma_\alpha\pi_\alpha=\pi_4+\gamma_5\tau\cdot\pi$.

We are not advocating the exploitation of this analogy as a practical way to make chiral-invariant Lagrangians. The existing method which uses nonlinear realizations directly is a simpler one to apply. However, there is another aspect of the analogy between π_α and p_α which leads to a formal development of some power. This lies in the notion of the *boost*. It was Wigner's discovery[10] that the momentum four-vector p_α could with great advantage be represented in the form

$$p_\alpha=(L_p)_{\alpha4}m, \quad (1.12)$$

where m denotes the rest mass $\sqrt{p^2}$, and $(L_p)_{\alpha\beta}$ denotes a 4×4 matrix belonging to the Lorentz group. The mass-shell constraint $p_\alpha p_\alpha=m^2$ is accounted for automatically in the representation (1.12) by the (pseudo) orthogonality of the matrix L_p. The power of this representation lies in the fact that it leads to the realization of the Lorentz group in terms of 3×3 orthogonal matrices (or 2×2 unitary and unimodular matrices). Thus, corresponding to the Lorentz transformation

$$p_\alpha\rightarrow p_\alpha'=\Lambda_{\alpha\beta}p_\beta, \quad (1.13)$$

one has the realization

$$\Lambda\rightarrow R(p,\Lambda)=L_{\Lambda p}^{-1}\Lambda L_p, \quad (1.14)$$

where $R(p,\Lambda)$ is, in effect, an ordinary space rotation. This realization is, of course, essentially the same as the nonlinear realization (1.5) obtained in this instance by setting $p_4=(\mathbf{p}^2+m^2)^{1/2}$. It is perhaps worth noticing that, insofar as finite-dimensional realizations are involved, the distinction between the compact group $SU(2)\times SU(2)$ and its noncompact relative $SL(2,C)$ is a minor one.

We propose to adopt the method of Wigner for dealing with nonlinear realizations. That is, we shall express the preferred field in the form

$$\pi_\alpha=(L_\pi)_{\alpha4}f, \quad (1.15)$$

where $(L_\pi)_{\alpha\beta}$ denotes a 4×4 matrix whose components are dynamical variables. These variables are not all independent. They are subject to the constraints

$$\sum_\beta(L_\pi)_{\alpha\beta}(L_\pi)_{\gamma\beta}=\delta_{\alpha\gamma}, \quad \det(L_\pi)=1 \quad (1.16)$$

or, in other words, L_π belongs to $SO(4)$. Included

[8] In a given model it might be possible to compute this quantity by a self-consistent method. That is, substitute $\chi=\langle\chi\rangle+\chi'$ in the Lagrangian, and set equal to zero the perturbation corrections to $\langle\chi'\rangle$.

[9] P. W. Anderson, Phys. Rev. 130, 439 (1963); P. W. Higgs, Phys. Letters 12, 132 (1964); Phys. Rev. 145, 1156 (1966); G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Letters 13, 585 (1964); T. W. B. Kibble, Ref. 7.

[10] E. P. Wigner, Ann. Math. 40, 149 (1939).

among the constraints (1.16) is of course the principal one, $\pi_\alpha\pi_\alpha = f^2$.

The main advantage to be gained by replacing the preferred field π_α with the matrix $(L_\pi)_{\alpha\beta}$ is the ease with which such a matrix can be used to effect a passage between linear and nonlinear realizations. Moreover, it enables one to discover the general features of a class of nonlinear realizations without the complication of having to commit oneself to a particular parametrization. It is this formal power which makes the boost approach useful for generalizing beyond the chiral groups. However, to avoid semantic confusion we shall invent the new name *reducing matrix* for the matrix L_π and its generalizations, since the word "boost" has already a rather precise meaning within the context of the inhomogeneous Lorentz group and its representations.

The realizations of a continuous group G, which become linear when restricted to some specified subgroup H, are treated by means of the reducing matrix in Sec. 2. These realizations are then gauged in the Yang-Mills manner. Section 3 contains some general remarks about symmetry breaking, both spontaneous and explicit. The formal techniques are illustrated in Sec. 4 on a model which could have practical interest, namely, the nonlinear realizations of $SU(3)$ which become linear with respect to $SU(2)_I \times U(1)_Y$.

2. NONLINEAR REALIZATIONS

Consider the problem of constructing nonlinear realizations of a continuous group G which become both linear and irreducible under some specified subgroup H. One may suppose that the linear irreducible representations of G,

$$\Psi \to D(g)\Psi, \quad g \in G \tag{2.1}$$

and their decomposition into linear irreducible representations of H are known. It will prove convenient to assume that the basis has been chosen so as to render the matrices $D(h)$, where $h \in H$, block-diagonal in form.

The first stage in solving the realization problem is the definition of a matrix $(L_\phi)_{\alpha\beta}$, the elements of which are field variables. This matrix, which we shall call the *reducing matrix*, will be subject to a number of algebraic constraints and will be endowed with a peculiar transformation behavior under the operations of the group G. The basic requirements are the following:

(a) The matrix L_ϕ is constrained to belong to the group G. i.e., to its self-representation. This is in order that, for any finite-dimensional representation $g \to D(g)$, the functional $D(L_\phi)$ shall be well defined. The number of independent fields, ϕ_a, needed to parametrize L_ϕ is therefore equal to or less than the dimensionality of G.

(b) Under the operations of the group G the fields which make up the reducing matrix transform according to

$$L_\phi \to g L_\phi h^{-1}(\phi,g), \tag{2.2}$$

where $g \in G$ and $h(\phi,g) \in H$. In other words, the columns of the reducing matrix must be arranged into sets which transform among themselves according to some representation of the subgroup H.

(c) Under the operations of the subgroup H the reducing matrix transforms in the ordinary way,

$$L_\phi \to h L_\phi h^{-1}. \tag{2.3}$$

It follows from (2.2) that the functionals $D(L_\phi)$, which are defined for any finite-dimensional representation, transform according to

$$D(L_\phi) \to D(g L_\phi h^{-1}) = D(g)D(L_\phi)D(h^{-1}). \tag{2.4}$$

It is this property which enables one to project nonlinear realizations out of linear ones such as (2.1) by the operation

$$\psi = D(L_\phi^{-1})\Psi. \tag{2.5}$$

A comparison of (2.1) and (2.4) yields for ψ the transformation law

$$\psi \to D(h)\psi, \tag{2.6}$$

where $h = h(\phi,g)$ is, in general, a nonlinear structure which depends upon the preferred fields ϕ_a which parametrize L_ϕ.

The detailed form of the matrix $h(\phi,g)$ is dependent upon the parametrization scheme, i.e., it depends upon which combinations of the components $(L_\phi)_{\alpha\beta}$ are taken as independent variables. Perhaps the simplest scheme is the one adopted by Coleman, Wess, and Zumino[8] and, earlier, by Kibble[7]:

$$L_\phi = e^{\phi \cdot A},$$

where A_a denotes the set of infinitesimal generators of G that are not contained in the algebra of H. Whatever the scheme chosen, one can discover the matrix $h(\phi,g)$ by referring Eq. (2.4) to a representation of G which contains a singlet of H. Such a representation includes at least one column, χ, for which (2.4) takes the form

$$D(L_\phi)\chi \to D(L_{\phi'})\chi = D(g)D(L_\phi)\chi, \tag{2.7}$$

i.e., for which $D(h)$ is represented by the identity. If the chosen parameters ϕ_a are expressed in terms of the components of the column which transforms according to (2.7), then it is straightforward to compute the transformation law of these parameters. Having done this, one can compute the matrix h by comparison with (2.2), i.e.,

$$h(\phi,g) = L_{\phi'}^{-1}\, g L_\phi. \tag{2.8}$$

The method will be illustrated in Paper II for the case of the conformal group.[11]

Consider now the problem of defining a *covariant derivative* operator for the nonlinear realization (2.6).

[11] A. Salam and J. Strathdee, following paper, Phys. Rev. 184, 1760 (1969).

It is evident that the ordinary derivative is not covariant:

$$\partial_\mu\psi \to D(h)\partial_\mu\psi+\partial_\mu D(h)\psi.$$

In order to be able to make covariant field equations it is essential that one define a covariant operator resembling the derivative. This can be done in the following way.

Let us imbed ψ in some linear representation $D(g)$:

$$\psi=D(L_\phi^{-1})\Psi;$$

and let us define, relative to it, the operator Δ_μ:

$$\Delta_\mu\psi=D(L_\phi^{-1})\partial_\mu\Psi, \tag{2.9}$$

which is clearly covariant. However, one should not adopt Δ_μ as the desired covariant derivative since it depends upon the imbedding representation $D(g)$. In order to remove this dependence it is necessary to analyze (2.9) more closely. Write

$$\Delta_\mu\psi=\partial_\mu\psi+D(L_\phi^{-1})\partial_\mu D(L_\phi)\psi. \tag{2.10}$$

The matrix $D^{-1}\partial_\mu D$ can be simplified if use is made of the constraint $L_\phi\in G$ for all x. In particular, it follows that the matrix

$$L_\phi^{-1}(x)L_\phi(x+\delta x)=1+\delta x_\mu L_\phi^{-1}\partial_\mu L_\phi+\cdots$$

is an infinitesimal transformation of G. In other words, the matrices $L_\phi^{-1}\partial_\mu L_\phi$ belong to the infinitesimal algebra of G. They can therefore be expanded in the form[11]

$$L_\phi^{-1}\partial_\mu L_\phi=i(L_\phi^{-1}\partial_\mu L_\phi)_i s_i, \tag{2.11}$$

where the matrices s_i constitute a basis of the algebra and the coefficients of the expansion are denoted by $(L_\phi^{-1}\partial_\mu L_\phi)_i$. In the representation $D(g)$, where the infinitesimal generators s_i are represented by S_i, the expansion (2.11) takes the form

$$D(L_\phi^{-1})\partial_\mu D(L_\phi)=i(L_\phi^{-1}\partial_\mu L_\phi)_i S_i. \tag{2.12}$$

The transformation behavior of the coefficients in (2.11) is complicated by the presence of the derivative operator. From (2.2) one finds that

$$L_\phi^{-1}\partial_\mu L_\phi \to h(L_\phi^{-1}\partial_\mu L_\phi)h^{-1}+h\partial_\mu h^{-1}. \tag{2.13}$$

That is, there is present, in general, an inhomogeneous term in the transformation law. Clearly, however, the inhomogeneity belongs to the algebra of H. This point is of crucial importance because it means that the fields $(L_\phi^{-1}\partial_\mu L_\phi)_i$ can be divided into two sets, one of which transforms covariantly while the other contains the inhomogeneity.

Let us suppose that the algebraic basis s_i has been chosen in such a way that it can split into two components m_α and n_a,[12] which transform independently under H. Moreover, suppose that the m_α constitute a basis for the subalgebra H. In this basis the expansion (2.11) takes the form

$$(1/i)L_\phi^{-1}\partial_\mu L_\phi=\Gamma_{\mu\alpha}m_\alpha+\lambda D_\mu\phi_a n_a, \tag{2.14}$$

which is to be looked upon as the definition of the field quantities $\Gamma_{\mu\alpha}$ and $D_\mu\phi_a$. The inhomogeneous term in the transformation (2.13) affects only the $\Gamma_{\mu\alpha}$; the fields $D_\mu\phi_a$ therefore belong to a *bona fide* nonlinear realization of the group G. *They are to be interpreted as the covariant derivatives of the preferred fields ϕ_a in terms of which the reducing matrix is parametrized.* The real parameter λ will be fixed by normalizing the kinetic-energy term associated with ϕ_a.

Corresponding to the expansion (2.14) one has, in the representation $D(g)$,

$$(1/i)D(L_\phi^{-1})\partial_\mu D(L_\phi)=\Gamma_{\mu\alpha}M_\alpha+\lambda D_\mu\phi_a N_a, \tag{2.15}$$

which can be substituted into the expression (2.10) for Δ_μ:

$$\Delta_\mu\psi=\partial_\mu\psi+i\Gamma_{\mu\alpha}M_\alpha\psi+i\lambda D_\mu\phi_a N_a\psi. \tag{2.16}$$

Since the left-hand side of (2.16) transforms covariantly, as does the third term on the right, therefore the sum of the remaining two terms must also be covariant. The latter part, denoted $D_\mu\psi$, has in addition the required property of being independent of the imbedding representation. Thus, the covariant derivative of ψ is given by

$$D_\mu\psi=\partial_\mu\psi+i\Gamma_{\mu\alpha}M_\alpha\psi. \tag{2.17}$$

Finally, consider the problems which arise when the transformations of the group G are made space-time–dependent; i.e., when G is turned into a gauge group of the Yang-Mills type, we have

$$\Psi(x) \to D(g)\Psi(x), \quad g=g(x)\in G. \tag{2.18}$$

There is no need to alter the prescription (2.5) for extracting the nonlinear realizations from Ψ. Indeed, the nonlinear transformation law (2.6) is formally unchanged since the matrix $h(\phi,g)$ is defined for arbitrary $g(x)\in G$. The modifications are needed, of course, in the definition of covariant derivatives.

Now, it is well known that the ordinary derivative is not covariant under space-time–dependent transformations:

$$\partial_\mu\Psi(x) \to D(g)\partial_\mu\Psi(x)+\partial_\mu D(g)\Psi(x). \tag{2.19}$$

With the basis s_i defined above, one can write

$$(1/i)g\partial_\mu g^{-1}=(g\partial_\mu g^{-1})_i s_i \tag{2.20}$$

since the matrices $g^{-1}\partial_\mu g$ belong to the algebra of G. Therefore, (2.19) can be expressed in the form

$$\partial_\mu\Psi(x) \to D(g)[\partial_\mu+i(g\partial_\mu g^{-1})_i S_i]\Psi(x). \tag{2.21}$$

In order to replace this with a covariant formula, one must introduce a set of gauge fields

$$A_\mu=A_{\mu i}s_i, \tag{2.22}$$

[11] A summation over the repeated subscript i is implied. On occasion it will be useful to divide the sum into two parts, one over the subalgebra H indicated by repeated greek subscripts $\alpha, \beta, \cdots$, and the remainder indicated by repeated latin subscripts from the beginning of the alphabet, $a, b, \cdots$.

which transform, according to the following law:

$$A_\mu \to gA_\mu g^{-1} + (1/if)g\partial_\mu g^{-1}. \tag{2.23}$$

The covariant derivative for linear representations is then defined by

$$\mathfrak{D}_\mu \Psi = (\partial_\mu + ifA_{\mu i}S_i)\Psi. \tag{2.24}$$

In the usual fashion, the covariant derivative of the gauge field itself is contained in the antisymmetric tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + if[A_\mu, A_\nu]. \tag{2.25}$$

From the expressions (2.24) and (2.25), which are covariant in the linear sense, one can project out the generalized nonlinear covariant derivatives. First, if ψ is defined by (2.5) its covariant derivative must be contained in the operator

$$\Delta_\mu \psi = D(L_\phi^{-1})(\partial_\mu + ifA_{\mu i}S_i)\Psi,$$

which can be simplified to the form

$$\Delta_\mu \psi = (\partial_\mu + ifB_{\mu i}S_i)\psi, \tag{2.26}$$

where $B_{\mu i}$ is a modified gauge field defined by

$$B_\mu = L_\phi^{-1}A_\mu L_\phi + (1/if)L_\phi^{-1}\partial_\mu L_\phi. \tag{2.27}$$

It transforms under a general gauge transformation $g(x)$ according to

$$B_\mu \to hB_\mu h^{-1} + (1/if)h\partial_\mu h^{-1}. \tag{2.28}$$

The most important feature of this transformation law is the fact that the inhomogeneous term belongs to the algebra of H. This means that the operator Δ_μ of (2.26) can be separated covariantly into two pieces,

$$\Delta_\mu \psi = D_\mu \psi + i\lambda(D_\mu \phi_a)N_a \psi,$$

which defines the generalized covariant derivatives

$$D_\mu \psi = (\partial_\mu + ifB_{\mu\alpha}M_\alpha)\psi \tag{2.29}$$

and

$$\lambda D_\mu \phi_a = fB_{\mu a}. \tag{2.30}$$

These expressions are covariant against gauge transformations of the second kind. It remains only to find the covariant derivatives of the fields $B_{\mu\alpha}$ and $B_{\mu a}$. These are contained in the antisymmetric tensor

$$B_{\mu\nu} = L_\phi^{-1}F_{\mu\nu}L_\phi,$$

which, in view of the definition (2.27), takes the form

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + if[B_\mu, B_\nu]. \tag{2.31}$$

The realizations discussed in this section can be used in the construction of Lagrangians that are invariant with respect to the transformations of G. First, a Lagrangian which is manifestly invariant with respect to the subgroup H can be modified so as to become invariant with respect to the *space-time-independent* transformations of the larger group G. It is necessary only to replace the ordinary derivatives $\partial_\mu \psi$ by their covariant form $D_\mu \psi$ as given in (2.17) and to take account of the new zero-spin boson field ϕ_a, whose existence this implies, by adding a covariant kinetic-energy term

$$\tfrac{1}{2}(D_\mu \phi_a^*)(D_\mu \phi_a),$$

and, possibly, adding other derivative coupling terms using $D_\mu \phi_a$. It is not possible to construct any covariant object which contains a term like $\phi_a^* \phi_a$, and it therefore follows that the new bosons must be without mass.

The invariance of the Lagrangian can be further enlarged to include the space-time-dependent transformations of G by introducing a gauge field B_μ which transforms according to the *reducible* nonlinear realization (2.28). The covariant derivatives $D_\mu \psi$ and $D_\mu \phi_a$ are given the new forms (2.29) and (2.30), respectively, while for the gauge field it is necessary to adjoin the kinetic-energy term

$$\tfrac{1}{4}B_{\mu\nu} \cdot B_{\mu\nu},$$

where $B_{\mu\nu}$ is given by (2.31). The upshot of these final modifications is that the preferred field ϕ_a and its massless quanta have disappeared from the Lagrangian. They have been absorbed by a redefinition (2.27) of the gauge field. Not only has the multiplet of massless bosons ϕ_a disappeared, but part of the gauge field, $B_{\mu a}$, has acquired a well-defined mass f/λ. The other part, $B_{\mu\alpha}$, which enters the covariant derivatives, remains without mass.

This phenomenon, whereby the introduction of a gauge multiplet of vector particles causes the disappearance of the massless zero-spin particles, has been discussed by a number of authors[9] in the context of spontaneous symmetry breaking. In Sec. 3 we show that this is precisely the context in which nonlinear realizations have meaning. The massless zero-spin particles are indeed the Goldstone bosons.

3. SPONTANEOUS SYMMETRY BREAKING

In this section we demonstrate that the formalism of nonlinear realizations and effective Lagrangians provides a natural framework for treating intrinsically broken symmetries. The arguments given here parallel those of Kibble.[7]

To discuss a system with spontaneous symmetry breaking one must have in mind a Lagrangian which is invariant with respect to the transformations of some continuous group G. Secondly, one must assume that the ground state or vacuum is not an invariant of G but only of some subgroup H. This property of the ground state is signalled by the nonvanishing expectation values of fields or combinations of fields which belong to nontrivial representations of G. Its consequences include symmetry-breaking perturbations of the masses and couplings of physical particles and, in particular, the appearance of spin-zero massless bosons.

Consider a system of fields, fermions and bosons, denoted collectively by Ψ and, in addition, a spin-zero multiplet M (of which some components M_a will correspond to Goldstone particles), which transform according to the reducible linear representations

$$\Psi \to D(g)\Psi, \quad M \to \hat{D}(g)M, \tag{3.1}$$

where g denotes an element of G, and D and $\hat{D}$ are the matrices appropriate to the representations concerned. The Lagrangian of this system is supposed to be invariant under these transformations:

$$L(\Psi,\partial_\mu\Psi,M,\partial_\mu M)=L(D\Psi,D\partial_\mu\Psi,\hat{D}M,\hat{D}\partial_\mu M). \tag{3.2}$$

This means that the system is classified into complete multiplets of G with the various couplings that are allowed by this symmetry. It may therefore be quite unlike the physical reality which reflects the ground-state asymmetry. The mass splittings of the physical multiplets can be large and, in fact, so large that some of the multiplets may be regarded as incomplete. The same is true for the couplings.

We know that in such a system Goldstone bosons must be present. Being massless, these particles cause a readjustment of the stable states of the system. In particular, the physical vacuum should contain an admixture of zero-energy Goldstone particles (which is to say that the vacuum degenerate). This property of the Goldstone particles can be formulated as effecting a redefinition of the "bare" masses and coupling constants that takes account of the ground-state asymmetry. We wish to introduce a set of fields $\phi_a(x)$ to represent the Goldstone particles and to put the Lagrangian (3.2) into a form which, *though still invariant under the transformations of* G, shows explicitly, in its bare masses and coupling constants, the effects of the underlying asymmetry. As stated in the Introduction, the method of nonlinear realizations, with the preferred fields $\phi_a(x)$, appears to be just the right construct to solve this problem.

The subset M_a of M referred to earlier transforms under the subgroup H, like the set of those generators n_a [see (2.14)] that correspond to the spontaneously broken symmetries of G. The remaining components of M, which are not included among the set M_a, will be labelled as M_A:

$$M=\begin{pmatrix} M_a \\ M_A \end{pmatrix}.$$

In such a case it is possible to invent a transformation $L_\phi(x)\in G$ which transforms away the components M_a in the sense that one can represent M in the form

$$M=\hat{D}(L_\phi)m, \tag{3.3}$$

with $L_\phi(x)$ so determined that

$$m_a=(\hat{D}(L_\phi^{-1})M)_a=0. \tag{3.4}$$

The components of the matrix L_ϕ—obtained as nonlinear functions of M by solving (3.4)—must satisfy various constraint conditions in order that L_ϕ belong to G, but they can be expressed in terms of a set of suitably chosen independent parameters ϕ_a, equal in number to the m_a of (3.4). In principle, therefore, one can solve for these parameters ϕ_a (the preferred fields) in terms of the original set of fields M. (The number of m_a's being set equal to zero equals the number of ϕ_a's introduced.) Since the latter fields transform according to the given linear rule (3.1), one can determine the transformation law of the preferred set ϕ_a. This law is quite generally nonlinear and, if the nonlinearities are expanded in powers, inhomogeneous. It therefore follows, as has been emphasized in Sec. 1, that the representation (3.3) can be used only in theories with noninvariant vacuums.[13]

It may be that some of the fields Ψ, and in particular M_A in (3.2), do not represent physical particles—in other words, they represent particles which are so far removed in mass from their partners in the set M as to be irrelevant dynamically [the analogy of M_a is with π_a $(a=1,2,3)$ and of M_A with σ in the chiral model]. The chief problem, therefore, is to exhibit the formalism in such a way that these fields can be removed from consideration. We do this by following the standard nonlinear prescription of imposing constraints; the details of the method are as follows.

In the Lagrangian (3.2) substitute the expression (3.3) for the fields M and its analog for Ψ to give

$$\begin{aligned} &L(\Psi,\partial_\mu\Psi,M,\partial_\mu M)\\ &\quad=L(D(L_\phi)\psi,\partial_\mu D(L_\phi)\psi,\hat{D}(L_\phi)m,\partial_\mu\hat{D}(L_\phi)m)\\ &\quad=L(\psi,D(L_\phi^{-1})\partial_\mu D(L_\phi)\psi,m,\hat{D}(L_\phi^{-1})\\ &\qquad\qquad\times\partial_\mu\hat{D}(L_\phi)m), \end{aligned} \tag{3.5}$$

where the validity of the second step depends upon the invariance of this Lagrangian under the transformations of G. The derivative terms in (3.5) involve the operator Δ_μ defined in Sec. 2,

$$D(L_\phi^{-1})\partial_\mu D(L_\phi)\psi=\Delta_\mu\psi=D_\mu\psi+i\lambda D_\mu\phi_a N_a\psi, \tag{3.6}$$

where $D_\mu\psi$ and $D_\mu\phi_a$ denote the covariant derivatives defined by (2.14) and (2.17). Similar relations hold for m. The fields ψ and m defined by (3.3) and (3.4) belong to a reducible nonlinear realization of G which becomes linear with respect to the subgroup H. The number of components ψ, ϕ_a, and m_A is equal to the number of original field components Ψ and M. However, it is clear that any subset of the fields ψ and m_A which spans a (linear) representation of H can be set equal to zero without doing violence to the invariance of L. Such a disappearance of some of the components of ψ is

[13] We do not have a general method which defines the transformation L_ϕ in cases where the Lagrangian contains explicitly no Goldstone boson fields with the requisite quantum numbers. For the above method to work, it is necessary to introduce extra fields with the requisite quantum numbers, perhaps by some self-consistent method which we have not investigated.

balanced by the appearance of algebraic constraints on Ψ. This can be seen by inverting the formulas (3.3) and (3.4) and expressing the components ψ and m_A as nonlinear functions[14] of Ψ and M.

Thus, one can express the Lagrangian (3.2) in terms of the nonlinear variables

$$L(\Psi,\partial_\mu\Psi,M,\partial_\mu M)=L(\psi,\Delta_\mu\psi,m_A,\Delta_\mu m_A), \quad (3.7)$$

and feed in the realistic bare masses and coupling constants. Unwanted components of ψ and m_A can now be set equal to zero[15]; the only rule to be observed is the manifest invariance of the right-hand side of (3.7) under the transformations of the subgroup H.

Contained in the Lagrangian (3.7) there will in general be the term

$$\tfrac{1}{2}D_\mu\phi_a D_\mu\phi_a{}^*=\tfrac{1}{2}\partial_\mu\phi_a\partial_\mu\phi_a{}^*+\text{interaction terms}.$$

The first term on the right-hand side of this is to be interpreted as the kinetic energy of the Goldstone bosons. These are massless bosons because, although ϕ_a appears elsewhere in the Lagrangian (in the covariant derivatives of ψ), it is always accompanied by $\partial_\mu\phi_a$. There is no mass term. That they are the Goldstone bosons is clear since they appear *only* when the symmetry is broken *spontaneously* by the imposition of constraints. It is clear from the above discussion that these particles could not even be defined if the vacuum were symmetric. On the other hand, if the Lagrangian were not symmetric they would acquire a mass since the matrix $D(L_\phi)$ would, in such a case, fail to be eliminated completely out of the right-hand side in (3.5).

It is of interest to see what happens if a long-range gauge field is present. Let us therefore replace the Lagrangian (3.2) by one which is invariant under transformations of the second kind. This means introducing a set of gauge fields, i.e.,

$$L(\Psi,\partial_\mu\Psi)\to L_1=L(\Psi,\partial_\mu\Psi+ifA_{\mu i}S_i\Psi)+\tfrac{1}{4}F_{\mu\nu i}F_{\mu\nu i}, \quad (3.8)$$

in the notation of Sec. 2. If the expression (3.3) for Ψ in terms of the nonlinear realization ψ is substituted in (3.8), one finds that

$$L_1=L(\psi,\partial_\mu\psi+ifB_{\mu i}S_i\psi)+\tfrac{1}{4}B_{\mu\nu i}B_{\mu\nu i}, \quad (3.9)$$

where B_μ is defined in terms of the gauge fields A_μ and the reducing matrix L_ϕ as in (2.27). B_μ transforms according to the nonlinear rule (2.28). The covariant derivative of B_μ is contained in the expression (2.31) for $B_{\mu\nu}$. The Goldstone particles, represented by L_ϕ, have been absorbed in the redefined gauge qelds B_μ. They no longer exist as independent particles. The fields ϕ_a still appear *implicitly* in the "nonlinear" transformation laws of ψ and B_μ, but they no longer have any dynamical significance. Moreover, the Lagrangian L_1, being independent of the ϕ_a, is not of the nonlinear variety.

Since the gauge fields $B_{\mu a}$ transform inhomogeneously [Eq. (2.28)], it is essential for the preservation of gauge invariance of the second kind that there should be no mass term $m^2B_{\mu a}{}^2$. However, it is possible to maintain gauge invariance of the first kind in the presence of a term such as

$$\tfrac{1}{2}m^2[B_{\mu a}-(1/if)(L_\phi{}^{-1}\partial_\mu L_\phi)_a]^2, \quad (3.10)$$

or, in other words, if the Goldstone particles are revived. This means that intrinsic symmetry breaking in the presence of a gauge field of *finite range* requires the presence of Goldstone particles.

In conclusion, the Goldstone fields ϕ_a will become massive if and only if there is introduced in (3.2) an explicit symmetry breaker. If that is done, it is clear that the reducing matrix L_ϕ must appear explicitly in the transformed Lagrangian (3.5) and not merely in the covariant derivatives. This means that ϕ_a is no longer everywhere accompanied by $\partial_\mu\phi_a$; thus, by expanding L_ϕ in powers of ϕ, one can always find a term proportional to ϕ^2. For example, one could add to the Lagrangian (3.5) an explicit symmetry breaker of the form

$$\kappa^2\hat{D}_{11}(L_\phi), \quad (3.11)$$

where $\hat{D}(g)$ denotes some chosen (self-conjugate) irreducible representation of G and $\hat{D}_{11}(g)$ indicates a matrix element of $\hat{D}(g)$ between states which are singlets of H, i.e.,

$$\hat{D}_{11}(hg)=\hat{D}_{11}(gh)=D_{11}(g). \quad (3.12)$$

The presence of a term like (3.11) in the Lagrangian therefore does not violate the symmetry under H.

This is the approach advocated by Weinberg.[16] A still more satisfactory Lagrangian, fully invariant under G but still producing a mass for the ϕ particles, could be

$$L=(L-\tfrac{1}{2}m^2\phi^2)+\tfrac{1}{2}m^2\phi^2,$$

where m^2 is computed self-consistently by setting up an interaction representation and computing the self-mass of the ϕ particle which is then put equal to the physical mass, i.e., its bare mass is zero. Whether this

[14] To illustrate, take the case of $M_a=\pi_a$, $a=1,2,3$, and $M_A=\sigma$. One can introduce (following Higgs and Kibble) a field $\chi(x)=\sqrt{(\pi^2+\sigma^2)}$. For the nonlinear method to apply, it is essential that $\chi_0=\langle\chi(x)\rangle$ should not vanish. Writing $\chi=\chi_0+\chi'$, where χ' is the quantized field, what we are saying in the text is that χ' can be set equal to zero.

[15] We have of course set the components $m_a=0$ by way of defining the preferred set ϕ_a. This does not reduce the number of independent variables, nor does it imply any constraints upon the M. When constraints are introduced, then m_A can also be set equal to zero so that all m's are out of (3.7). The Goldstone bosons, however, remain, being described by the ϕ_a's which appear in the definitions of Δ_μ. The formula (3.7) is analogous to Kibble's formula (10) (Ref. 7). Our formulation, being completely general, is valid for any symmetry group and any representation.

[16] Weinberg actually writes out the expression $\hat{D}_{11}(L_\phi)$ as a power series, and in his way of setting it out it is hard to recognize that it can be so compactly expressed.

self-consistency procedure will introduce other Goldstone particles into the theory is an open question.

4. A SIMPLE EXAMPLE

In order to illustrate the techniques presented in Sec. 2 we consider here the nonlinear realizations of $SU(3)$ which become linear with respect to the subgroup $SU(2)_I \times U(1)_Y$. This example is sufficiently complicated to exhibit the main features of the nonlinear formalism and, moreover, it has some physical relevance in that the breaking of $SU(3)$ symmetry may well be, at least in part, intrinsic. In addition, as can easily be seen, only very little effort will be needed to extend the formulas given here to the case of chiral $SU(3)\times SU(3)$ broken spontaneously to chiral $SU(2)\times SU(2)$.

The first stage in establishing the nonlinear realizations is the construction of a reducing matrix $(L_K)_\alpha{}^\beta \in SU(3)$. This matrix must transform according to

$$L_K \to g L_K h^{-1}(K,g), \tag{4.1}$$

where $g\in SU(3)$ and $h\in SU(2)\times U(1)$. Since $SU(2)\times U(1)$ is a four-parameter group, while $SU(3)$ has eight parameters, one expects that there should be a set of four preferred fields, K and $\bar{K}$, with which to parametrize L_K. These fields correspond to the hypercharge changing transformations of $SU(3)$. Out of all the possible parametrizations we shall pick one that does not involve square roots [and is therefore the nearest in spirit to Weinberg's treatment of chiral $SU(2)\times SU(2)$]. It is given by

$$L_K = \begin{pmatrix} \left(1-\dfrac{K\bar{K}}{\bar{K}K}\right)+\dfrac{1-\lambda^2\bar{K}K}{1+\lambda^2\bar{K}K}\dfrac{K\bar{K}}{\bar{K}K} & \dfrac{2\lambda K}{1+\lambda^2\bar{K}K} \\ \dfrac{2\lambda\bar{K}}{1+\lambda^2\bar{K}K} & \dfrac{1-\lambda^2\bar{K}K}{1+\lambda^2\bar{K}K} \end{pmatrix}, \tag{4.2}$$

where λ denotes a real parameter to be fixed later. The field K is a two-component column vector and $\bar{K}$ denotes its Hermitian adjoint, a two-component row vector.

It is necessary to demonstrate that the parametrization (4.2) is consistent with the transformation requirements (4.1). This can be done by setting up explicit transformation rules for the preferred fields K and $\bar{K}$. First, under the subgroup $SU(2)\times U(1)$, since $h(K,g)=g$, it is clear that K and $\bar{K}$ transform like $I=\frac{1}{2}$ fields with $Y=+1$ and $Y=-1$, respectively. To discover their behavior under the hypercharge changing transformations it is necessary to use the fact that $h^{-1}(K,g)$ multiplies the third column of L_K by a phase factor but does not mix into it the first two columns. Let us consider the infinitesimal hypercharge changing transformation

$$g=\begin{pmatrix} 1 & i\epsilon \\ i\bar{\epsilon} & 1 \end{pmatrix}, \tag{4.3}$$

where ϵ denotes a two-component infinitesimal quantity and $\bar{\epsilon}$ its Hermitian adjoint. The components of the column $(L_K)_\alpha{}^3$ are transformed according to

$$\delta\left(\frac{2\lambda K}{1+\lambda^2\bar{K}K}\right)=i\epsilon\frac{1-\lambda^2\bar{K}K}{1+\lambda^2\bar{K}K}+i\delta\varphi\frac{2\lambda K}{1+\lambda^2\bar{K}K},$$

$$\delta\left(\frac{1-\lambda^2\bar{K}K}{1+\lambda^2\bar{K}K}\right)=i\frac{2\lambda\bar{\epsilon}K}{1+\lambda^2\bar{K}K}+i\delta\varphi\frac{1-\lambda^2\bar{K}K}{1+\lambda^2\bar{K}K},$$

where $\delta\varphi=\delta\varphi(K,\epsilon)$ is a nonlinear effect coming from h^{-1}. It can be eliminated from these formulas, which then yield the transformation law

$$\delta\left(\frac{2\lambda K}{1-\lambda^2\bar{K}K}\right)=i\epsilon-\frac{2\lambda K}{1-\lambda^2\bar{K}K}\frac{2\lambda(i\bar{\epsilon}\cdot K)}{1-\lambda^2\bar{K}K}. \tag{4.4}$$

From this formula and its Hermitian adjoint it is a simple matter to extract the result

$$(1/i)\delta K=[(2\lambda)^{-1}(1-\lambda^2\bar{K}\cdot K)\epsilon-\lambda\bar{\epsilon}\cdot KK] -\tfrac{1}{2}\lambda\frac{1+\lambda^2\bar{K}K}{1-\lambda^2\bar{K}K}(\bar{\epsilon}\cdot K+\bar{K}\cdot\epsilon)K, \tag{4.5}$$

which is, therefore, the transformation law implied by (4.1) in the parametrization (4.2).

The infinitesimal form of the transformation $h(K,g)$ is determined by the method of Sec. 2,

$$1+\delta h=(L_K^{-1}+\delta L_K^{-1})(1+\delta g)L_K, \tag{4.6}$$

where δL_K is obtained by using (4.5) in conjunction with the form (4.2). The result is

$$(1/i)\delta h=\begin{pmatrix} -\lambda(K\bar{\epsilon}+\epsilon\bar{K})-\lambda^3\dfrac{K(\bar{K}\cdot\epsilon+\bar{\epsilon}\cdot K)\bar{K}}{1-\lambda^2\bar{K}K} & 0 \\ 0 & \lambda\dfrac{\bar{K}\cdot\epsilon+\bar{\epsilon}\cdot K}{1-\lambda^2\bar{K}K} \end{pmatrix}, \tag{4.7}$$

which clearly belongs to $SU(2)\times U(1)$. This matrix controls all of the nonlinear realizations with the exception of the preferred realization (4.5). Suppose that ψ transforms under $SU(2)\times U(1)$ according to a linear irreducible representation which is generated by the isospin and hypercharge matrices $\mathbf{I}$ and Y. These are defined, in the

context of $SU(3)$, by

$$\begin{aligned}\delta\psi=\delta h_\alpha{}^\beta F_\beta{}^\alpha\psi&=[\delta h_a{}^b(I_b{}^a+\tfrac{1}{3}\delta_b{}^a Y)-\delta h_3{}^3 Y]\psi\\&=[\delta h_a{}^b I_b{}^a+\tfrac{1}{2}(\delta h_1{}^1+\delta h_2{}^2-2\delta h_3{}^3)Y]\psi\\&=[\mathrm{tr}(\delta h\tau)\cdot\mathbf{I}-\tfrac{3}{2}\delta h_3{}^3 Y]\psi,\end{aligned}$$

where the $F_\alpha{}^\beta$ are defined on the $SU(3)$ quark by $F_\alpha{}^\beta\psi_\gamma=\delta_\gamma{}^\beta\psi_\alpha-\frac{1}{3}\delta_\alpha{}^\beta\psi_\gamma$. Then, corresponding to the hypercharge changing transformation (4.3) ,one finds, using (4.7) in (2.6),

$$\frac{1}{i}\delta\psi=-\lambda\left[(\bar\epsilon\tau K+\bar K\tau\epsilon)\cdot\mathbf{I}+\frac{\bar\epsilon\cdot K+\bar K\cdot\epsilon}{1-\lambda^2\bar K K}(\lambda^2\bar K\tau K\cdot\mathbf{I}+\tfrac{3}{2}Y)\right]\psi, \tag{4.8}$$

where τ denotes the Pauli matrices.

The covariant derivatives are determined in the parametrization (4.2) by the matrix

$$L_K^{-1}\partial_\mu L_K=\begin{bmatrix}\dfrac{2\lambda^2(K\bar K_\mu-K_\mu\bar K)}{1+\lambda^2\bar K K}-\dfrac{2\lambda^4 K(\bar K_\mu\cdot K-\bar K\cdot K_\mu)\bar K}{(1+\lambda^2 K\bar K)^2} & \dfrac{2\lambda K_\mu}{1+\lambda^2\bar K K}+2\lambda^3 K\dfrac{\bar K_\mu\cdot K-\bar K\cdot K_\mu}{(1+\lambda^2\bar K K)^2}\\ -\dfrac{2\lambda\bar K_\mu}{1+\lambda^2\bar K K}+2\lambda^2\dfrac{\bar K_\mu\cdot K-\bar K\cdot K_\mu}{(1+\lambda^2\bar K K)^2}\bar K & 2\lambda^2\dfrac{\bar K\cdot K_\mu-\bar K_\mu\cdot K}{(1+\lambda^2\bar K K)^2}\end{bmatrix}, \tag{4.9}$$

where $K_\mu=\partial_\mu K$ and $\bar K_\mu=\partial_\mu\bar K$. This matrix, being anti-Hermitian and traceless, belongs to the algebra of $SU(3)$. The part which belongs to the algebra of $SU(2)\times U(1)$ can be separated out and used in the construction of the covariant derivative of ψ,

$$\begin{aligned}D_\mu\psi=\partial_\mu\psi+2\lambda^2\Bigg[&\frac{\bar K_\mu\tau K-\bar K\tau K_\mu}{1+\lambda^2\bar K K}\cdot\mathbf{I}\\&-\frac{\bar K_\mu K-\bar K K_\mu}{(1+\lambda^2\bar K K)^2}(\lambda^2\bar K\tau K\cdot\mathbf{I}+\tfrac{3}{2}Y)\Bigg]\psi,\end{aligned} \tag{4.10}$$

and the remainder is used to define the covariant derivatives of K and $\bar K$:

$$\begin{aligned}D_\mu K&=\frac{K_\mu}{1+\lambda^2\bar K K}+\lambda^2 K\frac{\bar K_\mu K-\bar K K_\mu}{(1+\lambda^2\bar K K)^2},\\ D_\mu\bar K&=\frac{\bar K_\mu}{1+\lambda^2\bar K K}-\lambda^2\frac{\bar K_\mu K-\bar K K_\mu}{(1+\lambda^2\bar K K)^2}\bar K.\end{aligned} \tag{4.11}$$

Any Lagrangian made out of ψ and $\bar\psi$ and their covariant derivatives together with $D_\mu K$ and $D_\mu\bar K$ will be $SU(3)$-invariant if it conserves isospin and hypercharge. The vacuum state corresponding to such a Lagrangian will not be invariant under the hypercharge-changing transformations, and the resulting Goldstone bosons are characterized by the fields K and $\bar K$.

The interaction of, for example, nucleons N and hyerpons Λ with the Goldstone fields K and $\bar K$ could be described by the $SU(3)$-invariant Lagrangian

$$\begin{aligned}L=\bar N(i\gamma_\mu D_\mu-m_N)N+\bar\Lambda(i\gamma_\mu D_\mu-m_\Lambda)\Lambda+D_\mu\bar K D_\mu K\\+g_{\Lambda KN}\bar N\gamma_\mu\Lambda D_\mu K+\cdots,\end{aligned} \tag{4.12}$$

where $D_\mu\Lambda=\partial_\mu\Lambda$ and $D_\mu N$ is given by (4.10) with $\mathbf{I}=\tau/2$ and $Y=1$. To the invariant part (4.12) one might add an explicit symmetry-breaking term in order to give mass to K and $\bar K$. This term could take the form of a matrix element between $SU(2)\times U(1)$ singlets of the transformation $D(L_K)$ in the manner outlined in Sec. 3. The simplest representation D which can serve in this role is of course the octet. The octet symmetry breaking is given by

$$\begin{aligned}\frac{M^2}{12\lambda^2}\mathrm{tr}(L_K\lambda_8 L_K^{-1}\lambda_8)&=\frac{M^2}{6\lambda^2}\frac{1-4\lambda^2\bar K K+\lambda^4(\bar K K)^2}{1+2\lambda^2\bar K K+\lambda^4(\bar K K)^2}\\&=M^2/6\lambda^2-M^2\bar K K+2M^2\lambda^2(\bar K K)^2+\cdots,\end{aligned} \tag{4.13}$$

where the normalization has been adapted to give the kaons mass M. There is no necessity to stop at octet breaking. It would be a fairly straightforward calculation to derive the explicit form of a symmetry breaker belonging to any higher representation. In such calculations it is useful to define the field

$$\Phi=L_K\lambda_8 L_K^{-1}, \tag{4.14}$$

which transforms like an ordinary (linear) octet,

$$\Phi\to g\Phi g^{-1},\quad g\in SU(3)$$

but satisfies the algebraic constraints

$$\Phi(\Phi+1)=2. \tag{4.15}$$

The $I=Y=0$ component of Φ is proportional to the octet symmetry breaker (4.13). The generalized symmetry breakers, belonging to higher representations of $SU(3)$, could be expressed in terms of the $I=Y=0$ components of the appropriate irreducible tensor polynomials in Φ.

(Reprinted with permission from Nature, Vol. 252, No. 5484, pp. 569–571, December 13, 1974
© (1974) Macmillan Magazine Limited)

Transition to CP conservation and zero Cabibbo angle in strong magnetic fields

MOST current theories of particle symmetries assume that violations of these symmetries come about through a spontaneous breaking mechanism which produces non zero expectation values for certain scalar (elementary or composite) fields. As is well known (refs 1–7 and Lebedev Institute Reprint No. 101) these expectation values may make a phase transition to a zero value for certain critical temperatures and possibly also for certain critical external magnetic field strengths H_c, H_{c_1}, H_{c_2}, ... Here we point out that it is conceivable that the charge asymmetry (associated with CP violation) in $K_L \to \pi^{\pm} + l^{\pm} + \bar{\nu}(\nu)$ decays may disappear for fields of $\sim 8 \times 10^{10}$ gauss if CP violation is milli-weak in character, and that the Cabibbo angle may be reduced to zero—leading to suppression of certain hyperon decays—in fields of the order of 10^{16} gauss. These estimates are so strongly model-dependent that it may be worthwhile in any case, to make a systematic phenomenological search for effects on particle asymmetries of strong magnetic fields of 10^6 gauss upwards.

The notion of critical fields (above which spontaneously broken symmetries are restored) is well known from the Ginzburg–Landau theory of superconductivity which is also the prototype of spontaneous symmetry-breaking mechanisms used in particle physics. An order of magnitude estimate of the critical field H_c in this theory is provided by considering the free energy:

$$F = F_n + \alpha(T)|\varphi|^2 + \frac{\lambda(T)}{2}|\varphi|^4 + \ldots \tag{1}$$

where n denotes normal.

For temperature $T < T_c$, F has a minimum when

$$|\langle\varphi\rangle|^2 = -\alpha/\lambda \tag{2}$$

The 'thermodynamic' critical field close to T_c is then given by[8]:

$$\frac{H_c^2(T)}{8\pi} = F_n - F_s$$

$$= \frac{1}{2}\frac{\alpha^2}{\lambda}$$

$$= \frac{1}{2}\lambda|\langle\varphi\rangle|^4 \tag{3}$$

where s denotes superconducting.

Now in superconductivity theory the (Cooper-pair) field φ is itself charged and the magnetic field interacts directly with it. For K^0 fields (which we wish to investigate here) this is not the case. In an SU(3)-symmetric theory, however, the magnetic field does interact in higher order loop graphs. To see the qualitative emergence of a formula like equation (3), we later consider (for $T = 0$) an O(3)-symmetric gauge model, where the electromagnetic field is part of a gauge-triplet. Even though we are not dealing with a thermodynamic system, a critical field H_c seems to exist in a one-loop approximation above which the expectation value of the neutral field $\langle\varphi_3\rangle$ vanishes—the relationship between H_c and $\langle\varphi_3\rangle$ having the general form of equation (3). Here we shall use this formula for order of magnitude estimates. Clearly H_c will be small for those situations where $\langle\varphi\rangle$ and λ are small.

Now for strong interaction symmetries like SU(2) or SU(3), values of $\langle\varphi\rangle$ are typically $\sim$ BeV (or somewhat less), with $\lambda \simeq 1$. But for weak interaction symmetries—and these provide the more spectacular physical situations—with universal gauge coupling of order unity ($g^2 \approx 4\pi/137$, $g \approx 1/3$), $\langle\varphi\rangle$ values are typically tens or hundreds of BeV in magnitude, except for 'off-diagonal' situations. By way of illustration we consider the $\langle\varphi\rangle$ associated with real rotations of the n–Λ system (Cabibbo angle) and $\langle\varphi\rangle$ associated with complex rotations which signal CP violation in a milli-weak manner.

To estimate the magnitude of H_c for these cases, assuming that it exists we consider the ideas of Lee[9], Pais and Primack[10] or Mohapatra and Pati (ref. 11 and University of Maryland Technical Report No. 74–O85). In their milli-weak theory, the latter authors write the mass matrix for the n, Λ quark system in the form:

$$f(\bar{n},\bar{\Lambda})_L U_L \begin{bmatrix} m_n & 0 \\ 0 & m_\Lambda \end{bmatrix} V_R^{-1} \begin{bmatrix} n \\ \Lambda \end{bmatrix}_R \tag{4}$$

where U_L and V_R are matrices of the type:

$$\begin{vmatrix} \cos\theta_{L,R} & -\sin\theta_{L,R}\exp(i\delta_{L,R}) \\ \sin\theta_{L,R}\exp(i\delta_{L,R}) & \cos\theta_{L,R} \end{vmatrix} \tag{5}$$

with θ and δ real. The real part of the off-diagonal elements of the expectation-value matrix

$$U_L \begin{vmatrix} m_n & 0 \\ 0 & m_\Lambda \end{vmatrix} V_R^{-1}$$

(assuming for simplicity $\theta_L = \theta_R$) is given by:

$$\tfrac{1}{2}(m_\Lambda - m_N)\sin 2\theta_c \cos(\varepsilon/2) \tag{6}$$

where $\delta_L = \varepsilon/2$, $\delta_R = -\varepsilon/2$. Now from CP violation, ε in this model can be as small as 2×10^{-3}. With $m_\Lambda - m_N \simeq 175$ MeV, $\theta_c \sim 15°$ and assuming f in equation (4) to be typically $\simeq 1$, we obtain:

$$\mathrm{Re}\langle\varphi\rangle|_{\text{off-diagonal}} \simeq 45 \text{ MeV} \tag{7}$$

Using equation (3) and making the *ad hoc* assumption that the numerical constant (λ) appearing there is of order unity, we obtain

$$H_c \simeq 3 \times 10^{16} \text{ gauss} \tag{8}$$

(using the conversion $\langle\varphi\rangle$ (MeV) $= 0.3H$ (kgauss) cm).

If fields of this magnitude are applied, the real part of the 'off-diagonal' expectation value may vanish and, with it, the Cabibbo angle, suppressing certain hyperon and strange-particle decays.

Also, the imaginary part of the off-diagonal matrix element of the mass matrix is

$$(f/2)(m_\Lambda + m_N)\sin 2\theta_c \sin\varepsilon$$

This expectation value may make a transition to a zero value for fields $H_c \simeq 10^{14}$ gauss—and with it the (milli-weak) CP asymmetry—for a heavy quark model ($m_\Lambda \sim 5$ BeV); for a light quark model ($m_\Lambda \sim 300$ MeV), on the other hand, H_c may reduce to 8×10^{10} gauss.

(A preliminary and perhaps unrealistic model for spontaneous superweak CP breaking seems to give a value of around 3×10^6 gauss for the critical field strength. This is likely to be a very gross—though morale-building—underestimate, from the experimental point of view.)

Since in these estimates we have assumed, for simplicity, that all model-dependent constants like f and λ are $\sim$ unity and, further, since equation (3) itself is a very crude estimate, it is clear that H_c may change by orders of magnitude, and also not one but a hierarchy of critical fields may be discovered when a detailed investigation is carried through. This situation is not unfamiliar in superconductivity theory, where, depending on the ratio of the parameters λ and g (the gauge coupling), superconductors are divided into Type I and Type II, with Type II displaying a variety of critical fields[12] H_c, H_{c_1}, H_{c_2}, H_{c_3} and a corresponding variety of physical (vorticity) characteristics (For V_3Ga, for example, there are three widely differing critical fields, H_{c_1} $(T=0) \simeq 200$ gauss, H_c $(T=0) \sim 6{,}000$ gauss, and H_{c_2} $(T \simeq 0)$ which is as large as 300,000 gauss.)

Detailed models, in which we shall investigate the analogues of fields H_{c_1}, H_{c_2}, ... for particle physics, will be considered elsewhere.

Fields of strength $\sim 10^5$ gauss are possibly technically feasible for use in conjunction with K-beam experiments. Clearly, in order to test the ideas expressed here, stronger fields will be needed. There is no doubt, however, that if the basic ideas of spontaneous symmetry breaking are correct and if we believe in the validity of extrapolating from the one-loop calculation presented later, then there would exist critical fields for which the broken symmetries are likely to be restored. The future task of the theory is to explore such situations where the field strength required is not excessively intense and within reach of the forseeable technology. Since it is commonly assumed that the mean magnetic fields in pulsars are $\simeq 10^{12} \simeq 10^{14}$ gauss, it would seem—for what it is worth—that CP-violating phenomena do not take place in pulsars, though the Cabibbo angle is still non-zero.

To illustrate the effect of a uniform magnetic field on symmetry breaking, we give a simple model calculation. A Lagrangian which exhibits local O(3) symmetry is given by:

$$\mathscr{L} = \tfrac{1}{4}\mathbf{F}_{\mu\nu}{}^2 + \tfrac{1}{2}(\nabla_\mu \boldsymbol{\varphi})^2 + \frac{\mu^2}{2}\boldsymbol{\varphi}^2 - \frac{\lambda}{4}(\boldsymbol{\varphi}^2)^2$$

where the scalar and vector fields, $\boldsymbol{\varphi}$ and $\mathbf{A}_\mu$, are triplets with respect to the O(3) symmetry and the covariant derivatives which appear are

$$\nabla_\mu \boldsymbol{\varphi} = \partial_\mu \boldsymbol{\varphi} + e\mathbf{A}_\mu \times \boldsymbol{\varphi}$$
$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + 2e\mathbf{A}_\mu \times \mathbf{A}_\nu$$

The parameters μ^2 and λ are both positive: this implies a symmetry breakdown, O(3)—O(2), in the tree approximation. The vacuum expectation value of $\boldsymbol{\varphi}$ is non-vanishing, and indeed, may be used to define a direction in iso-space

$$\langle \varphi^i \rangle = \sqrt{(\mu^2/\lambda)}\delta^{i3}.$$

The corresponding component of the gauge potential, $A_\mu{}^3$, does not acquire a mass and so may be identified with the electromagnetic potential. In the tree approximation the presence of an external background electromagnetic field, $\langle A_\mu{}^3 \rangle \neq 0$, has no influence on the expected value of the neutral component $\langle \varphi^3 \rangle$ of the matter triplet.

When quantum corrections are taken into account this is no longer true. Already, the one-loop corrections to the effective action, given formally by:

$$\Gamma_{(1)}(\varphi, A) = (\hbar/2i)\,\mathrm{lnDet}G_{(0)}(\varphi, A)$$

(where $G_{(0)}$ denotes the set of classical propagation function for small disturbances on a given background) imply a coupling between φ^3 and $A_\mu{}^3$ even when the charged components $\varphi^{1,2}$ and $A_\mu{}^{1,2}$ are set equal to zero in the modified ground state.

In the uniform magnetic background, where all fields vanish except for

$$\varphi^3(x) = \psi, \quad A_1^3(x) = \tfrac{1}{2}Hx_2, \quad A_2^3(x) = -\tfrac{1}{2}Hx_1$$

with ψ and H constant, the one-loop contribution to the effective potential, $V_{(1)}(\psi, H)$, is determined approximately by the differential formula

$$2\frac{\partial V_{(1)}}{\partial \psi^2} = \frac{\lambda\hbar}{8\pi}\int_0^\infty \frac{\mathrm{d}s}{s}\left[\left\{\frac{eH}{\sinh(seH)}\exp\left[-s(\lambda\psi^2-\mu^2)\right] - \frac{\exp[-s(\lambda M^2-\mu^2)]}{s}\left(1-s\lambda(\psi^2-M^2)\right)\right\} + \frac{3}{2\pi s}\left\{\exp[-s(3\lambda\psi^2-\mu^2)] - \exp[-s(3\lambda M^2-\mu^2)] \times \left(1-3s\lambda(\psi^2-M^2)\right)\right\}\right]$$

(This expression is renormalised at $H=0$, $\psi=M$, that is, $\partial V_{(1)}/\partial\psi^2$ and $\partial^2 V_{(1)}/\partial(\psi^2)^2$ both vanish at this point.) The formula is approximate because we have not included the contributions due to intermediate vectors. The term with the H-dependent factor here is the contribution to $V_{(1)}$ of the intermediate charge pair $\varphi^{1,2}$, whereas the H-independent factor is due to the neutral φ^3.

To determine the critical field, we should set $\psi = 0$ and require that $2\partial V_{(1)}/\partial\psi^2|_{\psi=0}$ should compensate the tree contribution $2\partial V_{(0)}/\partial\psi^2|_{\psi=0} = -\mu^2$. Unfortunately, the neutral particle contribution to $\partial V^{(1)}/\partial\psi^2$ is complex if $\psi^2 < \mu^2/3\lambda$ This means that our approximation is inadequate. The source of this complexity is easily traced.

In restricting ourselves to one-loop effects we are not allowing the magnetic field to act on the neutral intermediate particles. Hence when $\psi = 0$ these appear to carry imaginary mass, $i\mu$. The very object of the computation $\partial V/\partial\psi^2 = 0$, is however, the setting to zero of the neutral particle mass. Clearly, we should be setting up a self-consistent (Dyson) equation: which means bringing in the contributions of higher loops. (For a similar pathology in a calculation of critical temperature see the paper of Dolan and Jackiw[5].) Short of this, we can obtain an order of magnitude idea of the critical field by setting both μ and ψ to zero specifically in the above expression for $\partial V_{(1)}/\partial\psi^2$. That is, we solve the restricted problem

$$0 = -\mu^2 + \frac{\lambda\hbar}{8\pi}\int_0^\infty \frac{\mathrm{d}s}{s}\left[\left\{\frac{eH_c}{\sinh(seH_c)} - \frac{\exp(-s\lambda M^2)}{s}(1+s\lambda M^2)\right\} + \frac{3}{2\pi s}\left\{1-\exp(-3s\lambda M^2)(1+3s\lambda M^2)\right\}\right]$$

or

$$\frac{\mu^2}{M^2} = \frac{\lambda^2\hbar}{8\pi}\int_0^\infty \frac{\mathrm{d}u}{u}\left[\frac{eH_c/\lambda M^2}{\sinh(ueH_c/\lambda M^2)} - \exp(-u)\left(\frac{1}{u}+1\right) + \frac{3}{2\pi}\left\{\frac{1}{u} - \exp(-3u)\left(\frac{1}{u}+3\right)\right\}\right] = \frac{\lambda^2\hbar}{8\pi}\left[a + b\frac{eH_c}{\lambda M^2}\right],$$

where

$$a = 1 + \frac{9}{2\pi} \simeq 2.5 \quad \text{and} \quad b = \int_0^\infty \frac{\mathrm{d}v}{v}\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{v}{\sinh v}\right) \simeq -1$$

On choosing the reference mass equal to μ we obtain

$$H_c = \frac{1}{b}\left(\frac{8\pi}{\lambda^{\frac{3}{4}}\hbar} - a\right)\frac{\lambda\mu^2}{e}$$

$$= \frac{1}{b}\left(\frac{8\pi}{\lambda^{\frac{3}{4}}\hbar} - a\right)\frac{\lambda^2\langle\varphi\rangle^2}{e}.$$

Since b is negative, we must have $\lambda^{\frac{3}{4}}\hbar > 8\pi/a$.

This estimate of H_c should not be taken as anything more than a rough indication. The form of this expression for H_c is basically similar to equation (3), apart from the numerical factor multiplying $\langle\varphi\rangle^2$. It is factors of this type which in a realistic calculation may drastically alter the order of magnitude estimates given earlier. Also, for a situation involving a number of non-zero $\langle\varphi\rangle$s, one may conjecture that H_c would be an expression involving a sum of terms like $\langle\Sigma f_i\varphi_i\rangle^2$ with sequences of opposing signs among the coefficients f_i. This, in turn, may lead to the desired diminution of the absolute magnitudes of the critical field strengths.

It is perhaps worth remarking that the calculation presented above for the critical field does not constitute a general proof that such fields always exist. This problem clearly needs further study.

We thank Professor D. J. Bradley for pointing out that fields of 10^9 gauss in a small volume $\sim 10^{-9}$ cm^3 lasting more than 10^{-9} s may conceivably be produced by laser compression techniques (though improvements in laser technology could permit the extension of both volume and duration) and to Dr Delbourgo for the remark that laser compression would also be accompanied by a high temperature which may help to produce the critical magnetic field at the same time. There is the further possibility that the electromagnetic energy pumping by the laser beam may by itself achieve the same purposes as are described in this communication. Such laser beams would have to have enormous powers of $\sim 10^{18}$ W cm^{-2}.

ABDUS SALAM

Imperial College,
Prince Consort Road,
London, SW7, UK
and
International Centre for Theoretical Physics

J. STRATHDEE

International Centre for Theoretical Physics,
Trieste, Italy

Received November 1, 1974.

[1] Brout, R., *Phase Transitions*, 5 (Benjamin, New York, 1965).
[2] Kirzhnits, D. A., *Soviet Phys. JETP*, **15**, 745 (1972).
[3] Kirzhnits, D. A., and Linde, A. D., *Phys. Lett.*, **42B**, 471 (1972).
[4] Weinberg, S., *Phys. Rev. D.*, **9**, 3357 (1974).
[5] Dolan, L., and Jackiw, R., *Phys. Rev. D.*, **9**, 3320 (1974).
[6] Bernard, C., *Phys. Rev. D.*, **9**, 3312 (1974).
[7] Harrington, B. J., and Yildiz, A., *Phys. Rev. Lett.*, **33**, 324 (1974).
[8] Saint James, D., Sarma, G., and Thomas, E. J., *Type II Superconductivity*, 23 (Pergamon, Oxford, 1969).
[9] Lee, T. D., *Phys. Rep.*, **9C**, No. 2 (1974).
[10] Pais, A., and Primack, J., *Phys. Rev. D.*, **8**, 3063 (1973).
[11] Mohapatra, R. N., and Pati, J. C., *Phys. Rev.*, **D8**, 2317 (1973).
[12] De Gennes, P. G., *Superconductivity of Metals and Alloys* (Benjamin, New York, 1966).

Reprinted with permission from *Nuclear Physics* **B96** (1975) 255–263
© 1975 North-Holland Publishing Company

INTENSITY-DEPENDENT MASS SHIFT AND SYMMETRY BREAKING

T.W.B. KIBBLE
Imperial College, London, England

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and
Imperial College, London, England

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 9 June 1975

We analyse the effects of the intensity-dependent mass shift predicted for electrons in an electromagnetic plane wave on bound states. It is shown that the null result of the experiment of Mowat et al. on the ^{133}Cs ground-state hyperfine splitting was to be expected, although their hypothetical explanation of it is incorrect. The recent suggestion that an intense laser beam might effect the restoration of a spontaneously broken symmetry is examined in more detail, and it is demonstrated that although a pure plane will not suffice, a superposition of plane waves, for example a standing wave, would do so. At present, however, the effect is unobservable small.

1. Introduction

Some years ago a mass shift of free electrons in an intense plane-wave electromagnetic field was predicted [1–3] and became the subject of lengthy controversy (summarized with full references in ref. [4]). Later it was suggested [5] that a similar effect might exist for bound electrons, leading to intensity-dependent shifts in the frequencies of spectral lines. However, an experimental search for such a shift in the hyperfine splitting of the ^{133}Cs ground state yielded negative results [6]. On a naive classical argument, this negative conclusion was to be expected, because the mass shift of a free electron can be understood [7] as a mean kinetic energy of the large amplitude oscillation induced by the field, which is precluded for bound electrons. The quantum-mechanical explanation, on the other hand, is less clear. Mowat et al. [6] suggested that the e^2A^2 term which is responsible should be regarded as a contribution to the energy rather than the mass, so yielding an unobservable uniform shift of all levels. However, this seems a rather unnatural

interpretation in view of the fact that in the Klein-Gordon (or second-order Dirac) equation the e^2A^2 term is clearly an addition to m^2. Tryon [8] in a treatment of quantum electrodynamics in a coherent photon background finds lowest-order energy level shifts independent of the mass shift, but neither the underlying reasons nor the extent of validity of this result are wholly clear.

In this paper we re-examine the status of the electron mass shift and show that, while the e^2A^2 term alone would yield level shifts corresponding to a change in mass rather than a uniform addition, the effect is almost exactly cancelled in the long wavelenth limit by the second-order shift arising from the $ep\cdot A$ term.

Recently [9] two of us suggested a different role for the mass shift, as a possible mechanism for the restoration of spontaneously broken symmetries, as we showed in the case of intense magnetic fields. Here we examine this possibility in more detail.

The mass shift directly affects only charged particles, whereas what is relevant in deciding whether a symmetry is broken is the effective potential evaluated as a function of the vacuum expectation value of some neutral field. An effect can be produced by virtual charged-particle loops but since the neutral particle is thus to be treated as a bound state, one might anticipate that the effect would be suppressed as in the case of an electron bound in an atom.

There is a simple invariance argument which shows that in a true plane wave the suppression is in fact complete. By gauge invariance, the effective potential $V(\phi, A_\mu)$ must depend on the plane-wave vector potential A_μ only through the field tensor $F_{\mu\nu}$. But it must also be a scalar, and for a plane wave no scalar can be formed from $F_{\mu\nu}$, so that V must be a function of the vacuum expectation value ϕ alone. This conclusion will be verified by more explicit examination of the effective potential. It follows that a spontaneously broken symmetry cannot be restored by a simple plane-wave laser beam, as has also been pointed out by Becker [10]. However the argument clearly fails for other configurations. In particular, a contribution to the effective potential can be expected from a superposition of two plane waves in different directions, for example from a standing wave produced by reflecting a laser beam from a mirror, since the invariant $F^{\mu\nu}F_{\mu\nu}$ is non-zero in that case.

It may also be worth remarking that, even in a pure plane-wave beam, symmetries might be broken (not restored) by a different mechanism, in which a symmetric tensor field acquires a vacuum expectation value proportional to $F^0_\mu F_{\rho\nu}$, or equivalently $A^2k_\mu k_\nu$.

It has been pointed out by Neville and Rohrlich [11] that a null-plane formalism is particularly well suited to the treatment of processes in a plane-wave field. In sect. 2 we introduce a version of this formalism closely based on the work of Bjorken, Kogut and Soper [12], in which the relativistic treatment closely parallels the non-relativistic. However we use a modified (non-orthogonal) co-ordinate system recently employed by Bell and Ruegg [13] in a treatment of the hydrogen atom. This method is interesting in its own right and may well have other applications.

In sect. 3 we show that the electron Green function in a plane-wave field has a particularly simple form in these coordinates, and use it to examine the contribution to an effective potential function. The conclusions are summarized and discussed in sect. 4.

2. Dirac equation in null-time coordinates

We use the system of coordinates

$$\tau = t - z, \qquad \zeta = z, \qquad x = (x, y), \tag{1}$$

comprising a null-time coordinate together with the three cartesian spatial coordinates. The distinction between ζ and z is required when we come to the contragredient transformations exemplified by the partial derivatives

$$\partial_\tau = \partial_t, \qquad \partial_\zeta = \partial_z + \partial_t, \qquad \partial = (\partial_x, \partial_y).$$

(In using the two-vector notation we regard all vectors other than ∂ as contravariant. Thus $p_\mu = i\partial_\mu$ but $\boldsymbol{p} = -i\partial$.) The metric tensor and its inverse in the coordinate system (1) are:

$$g_{\mu\nu} = \begin{bmatrix} 1 & 1 & . & . \\ 1 & . & . & . \\ . & . & -1 & . \\ . & . & . & -1 \end{bmatrix}, \qquad g^{\mu\nu} = \begin{bmatrix} . & 1 & . & . \\ 1 & -1 & . & . \\ . & . & -1 & . \\ . & . & . & -1 \end{bmatrix}.$$

Since γ^τ is singular, two of the four components of the Dirac equation take the form of constraints so that it can be reduced [12] to a two-component form which in our coordinates reads

$$2\pi_\tau \psi(x) = [\pi_\zeta + (m - i\sigma\cdot\pi)\pi_\zeta^{-1}(m + i\sigma\cdot\pi)]\,\psi(x). \tag{2}$$

Here

$$\pi_\mu = p_\mu - eA_\mu(x) = i\partial_\mu - eA_\mu(x), \tag{3}$$

and π_ζ^{-1} is the integral operator defined by

$$\pi_\zeta^{-1} f(x) = \frac{1}{2i}\int d\zeta'\, \epsilon(\zeta - \zeta') \exp\left[-ie\int_{\zeta'}^{\zeta} d\zeta''\, A_\zeta(\tau, \zeta'', x)\right] f(\tau, \zeta', x).$$

Let us first suppose that the external field is time-independent, $A_\mu = A_\mu(\zeta, x)$. Then eq. (2) has stationary solutions of the form

$$\psi(x) = u_n(\zeta, x)\, e^{-iE_n\tau}.$$

The u_n are eigenfunctions of the Hamiltonian

$$H = eA_\tau + \tfrac{1}{2}\pi_\zeta + \tfrac{1}{2}(m - i\sigma\cdot\pi)\pi_\zeta^{-1}(m + i\sigma\cdot\pi),$$

and are obtainable from the familiar Dirac wave functions by multiplying by $e^{-iE_n\zeta}$ and applying the projector $P_+ = \frac{1}{2}\gamma^\zeta\gamma^\tau$. The bound-state wave functions may be normalized so that

$$\int d\zeta\, d^2x\, u_n^\dagger(\zeta, x)\, u_{n'}(\zeta, x) = \delta_{nn'}. \tag{4}$$

The scattering states are labelled by the three-momentum p and z component of spin s. We use η to denote p_ζ, so that $p = (\eta, \boldsymbol{p})$. The covariant normalization of these states is

$$\int d\zeta\, d^2x\, u_{p's'}^\dagger(\zeta, x)\, u_{ps}(\zeta, x) = 2|\eta|\,(2\pi)^3\,\delta(\eta' - \eta)\,\delta_2(\boldsymbol{p}' - \boldsymbol{p})\,\delta_{s's}. \tag{5}$$

For a free particle,

$$u_{ps}(\zeta, x) = |2\eta|^{\frac{1}{2}}\, e^{i(p\cdot x - \eta\zeta)}\, w_s,$$

where w_s is a two-component spinor, $\binom{1}{0}$ or $\binom{0}{1}$. The corresponding energy eigenvalue is

$$E_p = \tfrac{1}{2}\eta + (m^2 + p^2)/2\eta. \tag{6}$$

For $\eta < 0$, u_{ps} is of course a negative-energy eigenfunction.

Now let us consider in addition to the time-independent field A_μ a plane wave field $a_\mu(\tau)$. Using standard perturbation methods, we look for a solution to the Dirac equation of the form:

$$\varphi(x) = \sum_n c_n(\tau)\, u_n(\zeta, x).$$

Choosing a gauge in which only the transverse components $\boldsymbol{a}$ of a_μ are non-zero, we find the equation

$$i\frac{dc_n(\tau)}{d\tau} = E_n\, c_n(\tau) + e\boldsymbol{a}(\tau)\cdot\sum_{n'} \boldsymbol{V}_{nn'}\, c_{n'}(\tau) + e^2\boldsymbol{a}^2(\tau)\sum_{n'} W_{nn'}\, c_{n'}(\tau), \tag{7}$$

which coincides in form with its non-relativistic counterpart, save for the definitions of $\boldsymbol{V}$ and W. These are

$$W_{nn'} = \frac{1}{2}\int d\zeta\, d^2x\, u_n^\dagger(\zeta, x)\, \pi_\zeta^{-1}\, u_{n'}(\zeta, x), \tag{8}$$

$$\boldsymbol{V}_{nn'} = -\frac{1}{2}\int d\zeta\, d^2x\, u_n^\dagger(\zeta, x)\,(\{\pi_\zeta^{-1}, \boldsymbol{\pi}\} + ie\sigma_3\,[\pi_\zeta^{-1}, {}^*\boldsymbol{\pi}])\, u_{n'}(\zeta, x), \tag{9}$$

where π_μ is still defined by (3) and ${}^*\boldsymbol{\pi}$ denotes the dual two-vector, ${}^*\pi^k = \epsilon^{kl}\,\pi^l$.

For a free particle both $\boldsymbol{V}$ and W are diagonal in the momentum representation. In fact,

$$\boldsymbol{V} = -\boldsymbol{p}/\eta, \qquad W = \tfrac{1}{2}\eta.$$

Hence we recover the solutions found earlier [14, 1, 3, 15] but in a simpler two-

component form,

$$c_{ps}(\tau) = c_{ps}(0) \exp\left(-i \int_0^\tau \mathrm{d}\tau' \, \hat{E}_p(\tau')\right),$$

with

$$\hat{E}_p(\tau) = E_p - \frac{ep\cdot a(\tau)}{\eta} + \frac{e^2 a^2(\tau)}{2\eta} = \tfrac{1}{2}\eta + \frac{[p - ea(\tau)]^2 + m^2}{2\eta}. \tag{10}$$

The absence of any explicit spin-dependent factor, similar to the one that appears in the four-component solution, is particularly interesting.

There is one crucial difference between (7) and the corresponding non-relativistic formula, namely that W as given by (8) is no longer a constant $(1/2m)$ but the matrix element of an operator $(1/2\eta_\zeta)$. Thus it yields an energy shift which is not constant but given approximately by

$$\Delta E_n = e^2 \langle a^2 \rangle_{\mathrm{av}} \, W_{nn} \approx e^2 \langle a^2 \rangle_{\mathrm{av}} \, E_n / 2m^2, \tag{11}$$

just the shift one would expect from a change in the electron mass.

If the frequency ω of the plane wave is small compared with a characteristic atomic frequency ω_0, then one may expect higher-order terms in the perturbation series to be negligible because of their large energy denominators. However, (11) cannot represent the true level shift, since for a wave of fixed intensity it tends to infinity in the long-wavelength limit. In fact, the matrix elements of V are of order $(\omega_0/m)^{\frac{1}{2}}$, so that the second-order term in V is of the same order of magnitude (e^2a^2/m) as the first-order term in W. That they actually cancel is seen most easily by making a gauge transformation of the type sometimes used to introduce the dipole approximation. This reduces a to zero while generating a new component

$$a'_\tau(\tau, x) = -x\cdot(\mathrm{d}a/\mathrm{d}\tau).$$

In this gauge the coefficients c'_n satisfy the equation

$$i\frac{\mathrm{d}c'_n(\tau)}{\mathrm{d}\tau} = E_n \, c'_n(\tau) - e\frac{\mathrm{d}a}{\mathrm{d}\tau}\cdot\sum_{n'} X_{nn'} \, c'_{n'}(\tau), \tag{12}$$

with

$$X_{nn'} = \int \mathrm{d}\zeta \, \mathrm{d}^2x \, u_n^\dagger(\zeta, x) \, x \, u_{n'}(\zeta, x). \tag{13}$$

This is an interesting formula because, although (12) looks like the familiar non-relativistic expression in the dipole approximation, it is in fact exact. [Recall that as compared with the usual integral, the integrand of (13) contains an extra factor of $e^{i(E_n - E_{n'})\zeta}$. This is just the standard exponential evaluated for an "energy-conserving" wave vector of magnitude $E_{n'} - E_n$ in the z direction.]

Since the matrix elements of X are of order $(m\omega_0)^{-\frac{1}{2}}$ and the diagonal matrix elements vanish, it is clear from (12) that the leading term in the energy-level shift is of order $(\omega/\omega_0)^2$ (e^2a^2/m), showing that the terms of order (e^2a^2/m) must can-

cel exactly. As $\omega \to 0$ we obtain precisely the usual second-order Stark shift, as we must.

3. Electron Green function and effective potentials

The electron Green function in a plane wave has been calculated by numerous authors [4]. Here we want to put in a particularly convenient form. In the two-component formalism, and using the notation (10), it may be written as

$$G(x, x') = \frac{i}{(2\pi)^3} \int d\eta\, d^2p\, [\theta(\eta)\,\theta(\tau - \tau') - \theta(-\eta)\,\theta(\tau' - \tau)]$$

$$\times \exp(-i\eta(\zeta - \zeta') + ip\cdot(x - x')) \exp\left(-i \int_{\tau'}^{\tau} d\tau''\, E_p(\tau'')\right). \tag{14}$$

Note that in common with the wave functions, it has no explicit spin dependence.

One can extract the gauge dependence as a line integral factor [16],

$$G(x, x') = e^{-ie\Lambda(x,x')}\, \hat{G}(x, x'), \tag{15}$$

with

$$\Lambda(x, x') = \int_{x'}^{x} dx^\mu\, a_\mu(x) = -(x - x')\cdot\langle a\rangle,$$

where the angular brackets denote a time average between τ' and τ,

$$\langle a\rangle = \frac{1}{\tau - \tau'} \int_{\tau'}^{\tau} d\tau''\, a(\tau'').$$

The gauge-invariant factor G may be written

$$G(x, x') = \int \frac{d^4p}{(2\pi)^4} \frac{e^{-ip(x-x')}\, 2\eta}{M^2(\tau, \tau') - p^2 - i\epsilon}, \tag{16}$$

where M^2 denotes the mass operator

$$M^2(\tau, \tau') = m^2 + e^2\langle a^2\rangle - e^2\langle a\rangle^2.$$

The non-Lorentz-invariant factor 2η in the integrand of (16) is required by the Lorentz transformation properties of $\psi(x)$. (Under longitudinal boosts, $\eta^{-\frac{1}{2}}$ is invariant.)

Because of the plane-wave character of the field, three components of energy momentum are conserved. In fact, the Fourier transform of (16) is

$$\hat{G}(p,p') = 2\eta(2\pi)^3\, \delta(\eta - \eta')\, \delta_2(p - p')$$

$$\times i \int d\tau\, e^{i(E-E')\tau} \int_0^\infty d\alpha \exp(-i\alpha[M^2(\tau + 2\alpha\eta, \tau) - p^2]). \tag{17}$$

A particularly simple and interesting case is that of a circularly polarized monochromatic wave

$$a(\tau) = \mathrm{Re}\,(\boldsymbol{a}e^{-i\omega\tau}), \qquad \boldsymbol{a}^2 = 0. \tag{18}$$

For this case a simple calculation shows that M^2 depends only on the difference of its arguments,

$$M^2(\tau, \tau') = m^2 + \Delta m^2 \left\{1 - \frac{\sin^2 [\frac{1}{2}\omega(\tau - \tau')]}{[\frac{1}{2}\omega(\tau - \tau')]^2}\right\} = M^2(\tau - \tau'), \tag{19}$$

where

$$\Delta m^2 = \tfrac{1}{2} e^2 \boldsymbol{a}\cdot\boldsymbol{a}^*$$

is the intensity-dependent mass shift. In this special case, G conserves all four components of energy momentum and can be written in a particularly compact form,

$$\hat{G}(p, p') = 2\eta\,(2\pi)^4\,\delta_4(p - p')\,i\int_0^\infty d\alpha \exp\,(-i\alpha[M^2(2\alpha\eta) - p^2]). \tag{20}$$

Physically, this means that the absorption and emission of photons form a circularly polarized plane wave is completely described by the time-integral factor in (15).

It is clear from (19) that while M^2 approaches the shifted mass $m^2 + \Delta m^2$ as $\tau - \tau' \to \infty$, for short time intervals it is approximaltely equal to the unshifted mass m^2. In particular $G(x, x)$ is independent of a_μ.

Now let us consider the possible effect of a plane-wave field a_μ on the effective potential function $V(\varphi)$ for some neutral scalar field φ. There is only one way to couple zero-momentum φ lines to a charged-fermion line, which amounts to replacing the fermion mass m by $m + g\varphi$. In particular, the one-loop contribution to $V(\varphi)$ is obtained from the single vacuum loop by this replacement. Since the vacuum loop involves $G(x, x)$, the one-loop potential is independent of $\boldsymbol{a}$.

As we noted in the introduction, this conclusion holds more generally, to all orders in the loop expansion, as a consequence of gauge and Lorentz invariance. However, it is true only for a plane-wave field, and only for vanishing external momenta. In other situations we may expect an effect.

Consider for example the one-loop contribution to the scalar field self-energy function π_φ (fig. 1), assuming the simplest invariant coupling of φ to the charged fermions. For the case of circular polarization (18), π_φ is translationally invariant and its Fourier transform is

Fig. 1. One-loop contribution to scalar-field self-energy.

$$\pi_\varphi(q) = \frac{ig^2}{(2\pi)^4}\int d^4p \int d^4p'\, \delta(p + p' - q)$$

$$\times \int_0^\infty d\alpha \exp\left(-i\alpha[M^2(2\alpha\eta) - p^2]\right) \int_0^\infty d\beta \exp\left(-i\beta[M^2(2\beta\eta') - p'^2]\right).$$

Using dimensional regularization to handle the quadratic divergence, the momentum integrals can be readily evaluated to give

$$\pi_\varphi(q) = -g^2 \int_0^\infty d\alpha \int_0^\infty d\beta\, [4\pi i(\alpha+\beta)]^{-\frac{1}{2}n}$$

$$\times \exp\left[i\frac{\alpha\beta}{\alpha+\beta}q^2 - i(\alpha+\beta)M^2\left(2\frac{\alpha\beta}{\alpha+\beta}q_\xi\right)\right]. \tag{21}$$

This formula exhibits π_φ as a function of the invariants q^2, $q\cdot k = q_\xi\omega$ and a^2. Clearly, the coefficient of a^2 vanishes when q or k tend to zero like $(q\cdot k)^2$, as it should since the combination $(q\cdot k)^2\, a^2$ is the gauge-invariant quantity $(F_{\mu\nu}\, q^\nu)^2$. For a neutral particle at rest, the energy shift induced by the plane wave is therefore smaller than the electron mass shift by a factor ω^2/m^2. Nevertheless, in suitable circumstances it might be observable.

Another case in which the external field could have an effect is that of a non-plane-wave field. The effective potential function $V(\varphi)$ must be a scalar function of $F_{\mu\nu}$, and can therefore have a non-trivial dependence in the external field when $F_{\mu\nu}F^{\mu\nu} \neq 0$. The contribution of a single plane wave vanishes essentially because there is only one wave vector k^μ involved, but when two or more such vectors appear this is no longer the case. In particular we may consider a superposition of plane waves travelling in opposite directions, described (in cartesian coordinates) by

$$k^\mu = (\omega, 0, 0, \omega), \qquad k'^\mu = (\omega, 0, 0, -\omega).$$

Physically, such a wave is readily produced by reflecting a laser beam from a mirror. Unfortunately, diagrams such as fig. 2 with only two external photon lines, and any number of zero-momentum φ lines, will give no contribution because of the delta function restricting the photons to have equal and opposite energy-momentum vectors. However, diagrams with four external photon lines will give a contribution

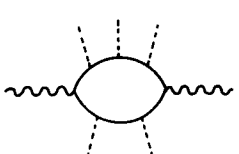


Fig. 2. Photon self-energy modified by any number of zero-momentum φ lines.

of order $e^2\omega^4 a^4$ to $V(\varphi)$. In the present circumstances this contribution is far too small to have observable consequences, but this may not always be true.

4. Conclusions

We have shown that the null result of the experimental search [6] for intensity-dependent level shifts in an atom in a plane-wave field was to be expected, despite the fact that the e^2A^2 interaction term is more properly regarded as generating a mass shift rather than a uniform energy shift. Indeed the shift is not uniform but is cancelled by second-order effects of the $ep\cdot A$ term.

Similarly we have shown that a pure plane-wave field will not induce restoration of spontaneously broken symmetries, because its contribution to the effective potential function vanishes for reasons of invariance. In principle, at least, a non-plane-wave field can have the desired effect. The simplest practical example is the standing wave produced by reflection from a mirror. However, at the present time the effect is far too small to be seen.

References

[1] N.D. Sengupta, Bull. Math. Soc. Calcutta 44 (1952) 175.
[2] A.I. Nikishov and V.I. Ritus, ZETF (USSR) 46 (1963) 776 [English transl.: JETP (Sov. Phys.) 19 (1964) 529].
[3] L.S. Brown and T.W.B. Kibble, Phys. Rev. 133 (1964) A705.
[4] J.H. Eberly, Progress in optics, ed. E. Wolf (North-Holland, Amsterdam, 1969) vol. 7.
[5] H.R. Reiss, Phys. Rev. Letters 17 (1966) 1162.
[6] J.R. Mowat, C.E. Johnson, H.A. Shugart and V.J. Ehlers, Phys. Rev. D3 (1971) 43.
[7] T.W.B. Kibble, Phys. Rev. 150 (1966) 1060.
[8] E.P. Tryon, Phys. Rev. Letters 32 (1974) 1139.
[9] A. Salam and J. Strathdee, Nucl. Phys. B90 (1975) 203.
[10] W. Becker, Private communication.
[11] R.A. Neville and F. Rohrlich, Nuovo Cimento 1A (1971) 625.
[12] J.D. Bjorken, J.B. Kogut and D.W. Soper, Phys. Rev. D3 (1971) 1382.
[13] J.S. Bell and H. Ruegg, Nucl. Phys. B93 (1975) 12.
[14] D.M. Volkov, Z. Phys. 94 (1935) 250.
[15] I.I. Goldman, Phys. Letters 8 (1964) 103.
[16] Ref. [3], appendix A.

Gauge unification of fundamental forces*

Abdus Salam

Department of Physics, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BZ England
and International Centre for Theoretical Physics, Prosecco-Miramere 21, 34100 Trieste, Italy

CONTENTS

I. Fundamental Particles, Fundamental Forces, and Gauge Unification 526
II. The Emergence of Spontaneously Broken SU(2) ×U(1) Gauge Theory 526
III. The Present and Its Problems 530
IV. Direct Extrapolation from the Electroweak to the Electronuclear 531
A. The three ideas 531
B. Tests of electronuclear grand unification 532
V. Elementarity: Unification with Gravity and Nature of Charge 533
A. The quest for elementarity, prequarks (preons and pre-preons) 533
B. Post-Planck physics, supergravity, and Einstein's dreams 534
C. Extended supergravity, SU(8) preons, and composite gauge fields 535
Appendix A: Examples of Grand Unifying Groups 536
Appendix B: Does the Grand Plateau really exist 536
References 537

In June 1938, Sir George Thomson, then Professor of Physics at Imperial College, London, delivered his 1937 Nobel Lecture. Speaking of Alfred Nobel, he said: "The idealism which permeated his character led him to ... (being) as much concerned with helping science as a whole, as individual scientists. ... The Swedish people under the leadership of the Royal Family and through the medium of the Royal Academy of Sciences have made Nobel Prizes one of the chief causes of the growth of the prestige of science in the eyes of the world... As a recipient of Nobel's generosity, I owe sincerest thanks to them as well as to him."

I am sure I am echoing my colleagues' feelings as well as my own, in reinforcing what Sir George Thomson said—in respect to Nobel's generosity and its influence on the growth of the prestige of science. Nowhere is this more true than in the developing world. And it is in this context that I have been encouraged by the Permanent Secretary of the Academy—Professor Carl Gustaf Bernhard—to say a few words before I turn to the scientific part of my lecture.

Scientific thought and its creation is the common and shared heritage of mankind. In this respect, the history of science, like the history of all civilization, has gone through cycles. Perhaps I can illustrate this with an actual example.

Seven hundred and sixty years ago, a young Scotsman left his native glens to travel south to Toledo in Spain. His name was Michael, his goal to live and work at the Arab Universities of Toledo and Cordova, where the greatest of Jewish scholars, Moses bin Maimoun, had taught a generation before.

Michael reached Toledo in 1217 AD. Once in Toledo, Michael formed the ambitious project of introducing Aristotle to Latin Europe, translating not from the original Greek, which he knew not, but from the Arabic translation then taught in Spain. From Toledo, Michael traveled to Sicily, to the Court of Emperor Frederick II.

Visiting the medical school at Salerno, chartered by Frederick in 1231, Michael met the Danish physician, Henrik Harpestraeng—later to become Court Physician of Eric IV Waldemarssön. Henrick had come to Salerno to compose his treatise on blood-letting and surgery. Henrik's sources were the medical canons of the great clinicians of Islam, Al-Razi and Avicenna, which only Michael the Scot could translate for him.

Toledo's and Salerno's schools, representing as they did the finest synthesis of Arabic, Greek, Latin, and Hebrew scholarship, were some of the most memorable of international assays in scientific collaboration. To Toledo and Salerno came scholars not only from the rich countries of the East, like Syria, Egypt, Iran and Afghanistan, but also from developing lands of the West like Scotland and Scandinavia. Then, as now, there were obstacles to this international scientific concourse, with an economic and intellectual disparity between different parts of the world. Men like Michael the Scot or Henrik Harpestraeng were singularities. They did not represent any flourishing schools of research in their own countries. With all the best will in the world their teachers at Toledo and Salerno doubted the wisdom and value of training them for advanced scientific research. At least one of his masters counseled young Michael the Scot to go back to clipping sheep and to the weaving of woolen cloth.

In respect of this cycle of scientific disparity, perhaps I can be more quantitative. George Sarton, in his monumental five-volume *A History of Science*, chose to divide his story of achievement in sciences into ages, each age lasting half a century. With each half century he associated one central figure. Thus 450 BC–400 BC Sarton calls the Age of Plato; this is followed by half centuries of Aristotle, of Euclid, of Archimedes, and so on. From 600 AD to 650 AD is the Chinese half century of Hsiian Tsang, from 650 to 700 AD that of I-Ching, and then from 750 AD to 1100 AD—350 years continuously—it is the unbroken succession of the Ages of Jabir, Khwarizmi, Razi, Masudi, Wafa, Biruni, and Avicenna, and then Omar Khayam—Arabs, Turks, Afghans, and Persians. After 1100 appear the first Western names: Gerard of Cremona, Roger Bacon—but the honors are still shared with the names of Ibn-

*This lecture was delivered December 8, 1979, on the occasion of the presentation of the 1979 Nobel Prizes in Physics.

Copyright © 1980 The Nobel Foundation

Rushd (Averroes), Moses Bin Maimoun, Tusi, and Ibn-Nafis—the man who anticipated Harvey's theory of circulation of blood. No Sarton has yet chronicled the history of scientific creativity among the pre-Spanish Mayas and Aztecs, with their re-invention of the zero, of the calendars of the moon and Venus and of their diverse pharmacological discoveries, including quinine, but the outline of the story is the same—one of undoubted superiority to the Western contemporary correlates.

After 1350, however, the developing world loses out except for the occasional flash of scientific work, like that of Ulugh Beg—the grandson of Timurlane, in Samarkand in 1400 AD; or of Maharaja Jai Singh of Jaipur in 1720—who corrected the serious errors of the then Western tables of eclipses of the sun and the moon by as much as six minutes of arc. As it was, Jai Singh's techniques were surpassed soon after with the development of the telescope in Europe. As a contemporary Indian chronicler wrote: "With him on the funeral pyre, expired also all science in the East." And this brings us to this century when the cycle begun by Michael the Scot turns full circle, and it is we in the developing world who turn westward for science. As Al-Kindi wrote 1100 years ago: "It is fitting then for us not to be ashamed to acknowledge truth and to assimilate it from whatever source it comes to us. For him who scales the truth there is nothing of higher value than truth itself; it never cheapens nor abases him."

Ladies and Gentlemen, it is in the spirit of Al-Kindi that I start my lecture with a sincere expression of gratitude to the modern equivalents of the Universities of Toledo and Cordova, which I have been privileged to be associated with—Cambridge, Imperial College, and the Centre at Trieste.

I. FUNDAMENTAL PARTICLES, FUNDAMENTAL FORCES, AND GAUGE UNIFICATION

The Nobel lectures this year are concerned with a set of ideas relevant to the gauge unification of the electromagnetic force with the weak nuclear force. These lectures coincide nearly with the 100th anniversary of the death of Maxwell, with whom the first unification of forces (electric with the magnetic) matured and with whom gauge theories originated. They also nearly coincide with the 100th anniversary of the birth of Einstein—the man who gave us the vision of an ultimate unification of *all* forces.

The ideas of today started more than twenty years ago, as gleams in several theoretical eyes. They were brought to predictive maturity over a decade back. And they started to receive experimental confirmation some six years ago.

In some senses then, our story has a fairly long background in the past. In this lecture I wish to examine some of the theoretical gleams of today and ask the question if these may be the ideas to watch for maturity twenty years from now.

From time immemorial, man has desired to comprehend the complexity of nature in terms of as few elementary concepts as possible. Among his quests—in Feynman's words—has been the one for "wheels within wheels"—the task of natural philosophy being to discover the innermost wheels if any such exist. A second quest has concerned itself with the fundamental forces which make the wheels go round and enmesh with one another. The greatness of gauge ideas—of gauge field theories—is that they reduce these two quests to just one; elementary particles (described by relativistic quantum fields) are representations of certain charge operators, corresponding to gravitational mass, spin, flavor, color, electric charge, and the like, while the fundamental forces are the forces of attraction or repulsion between these same *charges*. A third quest seeks for a *unification* between the charges (and thus of the forces) by searching for a single entity, of which the various charges are components in the sense that they can be transformed one into the other.

But are all fundamental forces gauge forces? Can they be understood as such, in terms of charges—and their corresponding currents—only? And if they are, how many charges? What unified entity are the charges components of? What is the nature of charge? Just as Einstein comprehended the nature of gravitational charge in terms of space-time curvature, can we comprehend the nature of the other charges—the nature of the entire unified set, *as a set*, in terms of something equally profound? This briefly is the dream, much reinforced by the verification of gauge theory predictions. But before I examine the new theoretical ideas on offer for the future in this particular context, I would like your indulgence to range over a one-man, purely subjective, perspective in respect of the developments of the last twenty years themselves. The point I wish to emphasize during this part of my talk was well made by G. P. Thomson in his 1937 Nobel Lecture. G. P. said "... The goddess of learning is fabled to have sprung full grown from the brain of Zeus, but it is seldom that a scientific conception is born in its final form, or owns a single parent. More often it is the product of a series of minds, each in turn modifying the ideas of those that came before, and providing material for those that come after."

II. THE EMERGENCE OF SPONTANEOUSLY BROKEN SU(2) × U(1) GAUGE THEORY

I started physics research thirty years ago as an experimental physicist in the Cavendish, experimenting with tritium–deuterium scattering. Soon I knew the craft of experimental physics was beyond me—it was the sublime quality of patience—patience in accumulating data, patience with recalcitrant equipment—which I sadly lacked. Reluctantly I turned my papers in, and started instead on quantum field theory with Nicholas Kemmer in the exciting department of P.A.M. Dirac.

The year 1949 was the culminating year of the Tomonaga–Schwinger–Dyson reformulation of renormalized Maxwell–Dirac gauge theory, and its triumphant experimental vindication. A field theory must be renormalizable and be capable of being made free of infinities—first discussed by Waller—if perturbative calculations with it are to make any sense. More—a renormalizable theory, with no dimensional parameter in its interaction term, connotes *somehow* that the fields represent "structureless" elementary entities.

With Paul Matthews, we started on an exploration of renormalizability of meson theories. Finding that renormalizability held only for spin-zero mesons and that these were the only mesons that empirically existed then, (pseudoscalar pions, invented by Kemmer, following Yukawa) one felt thrillingly euphoric that with the triplet of pions (considered as the carriers of the strong nuclear force between the proton–neutron doublet) one might resolve the dilemma of the origin of this particular force. By the same token, the so-called weak nuclear force—the force responsible for β radioactivity (and described then by Fermi's nonrenormalizable theory) had to be mediated by some unknown spin-zero mesons if it was to be renormalizable. If massive charged spin-one mesons were to mediate this interaction, the theory would be nonrenormalizable, according to the ideas then.

Now this agreeably renormalizable spin-zero theory for the pion was a field theory, but not a gauge field theory. There was no conserved charge which determined the pionic interaction. As is well known, shortly after the theory was elaborated, it was found wanting. The $(\frac{3}{2}, \frac{3}{2})$ resonance Δ effectively killed it off as a fundamental theory; we were dealing with a complex dynamical system, not "structureless" in the field-theoretic sense.

For me, personally, the trek to gauge theories as candidates for fundamental physical theories started in earnest in September 1956—the year I heard at the Seattle Conference, Professor Yang expound his and Professor Lee's ideas (Lee and Yang, 1956) on the possibility of the hitherto sacred principle of left–right symmetry, being violated in the realm of the *weak nuclear force*. Lee and Yang had been led to consider abandoning left–right symmetry for weak nuclear interactions as a possible resolution of the (τ, θ) puzzle. I remember traveling back to London on an American Air Force (MATS) transport flight. Although I had been granted, for that night, the status of a Brigadier or a Field Marshal—I don't quite remember which—the plane was very uncomfortable, full of crying servicemen's children—that is, the children were crying, not the servicemen. I could not sleep. I kept reflecting on why Nature should violate left–right symmetry in weak interactions. Now the hallmark of most weak interactions was the involvement in radioactivity phenomena of Pauli's neutrino. While crossing over the Atlantic, came back to me a deeply perceptive question about the neutrino which Professor Rudolf Peierls had asked when he was examining me for a Ph.D. a few years before. Peierls' question was: "The photon mass is zero because of Maxwell's principle of a gauge symmetry for electromagnetism; tell me, why is the neutrino mass zero?" I had then felt somewhat uncomfortable at Peierls, asking for a Ph.D. viva, a question of which he himself said he did not know the answer. But during that comfortless night the answer came. The analog for the neutrino of the gauge symmetry for the photon existed: it had to do with the masslessness of the neutrino, with symmetry under the γ_5 transformation (Salam, 1957a) (later christened "chiral symmetry"). The existence of this symmetry for the massless neutrino must imply a combination $(1+\gamma_5)$ or $(1-\gamma_5)$ for the neutrino interactions. Nature had the choice of an aesthetically satisfying but a left–right symmetry violating theory, with a neutrino which travels exactly with the velocity of light; or alternatively a theory where left–right symmetry is preserved, but the neutrino has a tiny mass—some ten thousand times smaller than the mass of the electron.

It appeared at that time clear to me what choice Nature must have made. Surely, left–right symmetry must be sacrificed in all neutrino interactions. I got off the plane the next morning, naturally very elated. I rushed to the Cavendish, worked out the Michel parameter and a few other consequences of γ_5 symmetry, rushed out again, got onto a train to Birmingham where Peierls lived. To Peierls I presented my idea: he had asked the original question; could he approve of the answer? Peierls' reply was kind but firm. He said "I do not believe left–right symmetry is violated in weak nuclear forces at all." Thus rebuffed in Birmingham, like Zuleika Dobson, I wondered where I could go next and the obvious place was CERN in Geneva, with Pauli—the father of the neutrino—nearby in Zurich. At that time CERN lived in a wooden hut just outside Geneva airport. Besides my friends, Prentki and d'Espagnat, the hut contained a gas ring on which was cooked the staple diet of CERN—Entrecôte à la creme. The hut also contained Professor Villars of MIT, who was visiting Pauli the same day in Zurich. I gave him my paper. He returned the next day with a message from the Oracle: "Give my regards to my friend Salam and tell him to think of something better." This was discouraging, but I was compensated by Pauli's excessive kindness a few months later, when Mrs. Wu's (Wu *et al.*, 1957), Lederman's (Garwin *et al.*, 1957) and Telegdi's (Friedman and Telegdi, 1957) experiments were announced showing that left–right symmetry was indeed violated and ideas similar to mine about chiral symmetry were expressed independently by Landau (1957) and Lee and Yang (1957). I received Pauli's first, somewhat apologetic letter on 24 January 1957. Thinking that Pauli's spirit should by now be suitably crushed, I sent him two short notes (Salam, 1957b)[1] I had written in the meantime. These contained suggestions to extend chiral symmetry to electrons and muons, assuming that their masses were a consequence of what has come to be known as dynamical spontaneous symmetry breaking. With chiral symmetry for electrons, muons, and neutrinos, the only mesons that could mediate weak decays of the muons would have to carry spin one. Reviving thus the notion of charged intermediate *spin-one* bosons, one could then postulate for these a type of gauge invariance which I called the "neutrino gauge." Pauli's reaction was swift and terrible. He wrote on 30th January 1957, then on 18 February and later on 11, 12, and 13 March: "I am reading (along the shores of Lake Zurich) in bright sunshine quietly your paper ..." "I am very much startled on the title of your paper 'Universal Fermi Interaction' ... For quite a while I have for myself the rule if a theoretician says *universal* it just means pure nonsense. This holds particularly in connection with

[1]For reference, see Footnote 7, p. 89, of Marshak, Riazuddin, and Ryan (1969), and W. Pauli's letters (CERN Archives).

the Fermi interaction, but otherwise too, and now you too, Brutus, my son, come with this word...." Earlier, on 30 January, he had written "There is a similarity between this type of gauge invariance and that which was published by Yang and Mills... In the latter, of course, no γ_5 was used in the exponent." and he gave me the full reference of Yang and Mills' paper, [Phys. Rev. 96, 191 (1954)]. I quote from his letter: "However, there are dark points in your paper regarding the vector field B_μ. If the rest mass is infinite (or very large), how can this be compatible with the gauge transformation $B_\mu - B_\mu - \delta_\mu \Lambda$?" and he concludes his letter with the remark: "Every reader will realize that you deliberately conceal here something and will ask you the same questions." Although he signed himself "With friendly regards," Pauli had forgotten his earlier penitence. He was clearly and rightly on the warpath.

Now the fact that I was using gauge ideas similar to the Yang–Mills [non-Abelian SU(2)-invariant] gauge theory was no news to me. This was because the Yang–Mills theory (Yang and Mills, 1954) (which married gauge ideas of Maxwell with the internal symmetry SU(2) of which the proton–neutron system constituted a doublet) had been independently invented by a Ph.D. pupil of mine, Ronald Shaw (1955), at Cambridge at the same time as Yang and Mills had written. Shaw's work is relatively unknown; it remains buried in his Cambridge thesis. I must admit I was taken aback by Pauli's fierce prejudice against universalism—against what we would today call unification of basic forces—but I did not take this too seriously. I felt this was a legacy of the exasperation which Pauli had always felt at Einstein's somewhat formalistic attempts at unifying gravity with electromagnetism—forces which in Pauli's phrase "cannot be joined—for God hath rent them asunder." But Pauli was absolutely right in accusing me of darkness about the problem of the masses of the Yang–Mills fields; one could not obtain a mass without wantonly destroying the gauge symmetry one had started with. And this was particularly serious in this context, because Yang and Mills had conjectured the desirable renormalizability of their theory with a proof which relied heavily and exceptionally on the masslessness of their spin-one intermediate mesons. The problem was to be solved only seven years later with the understanding of what is now known as the Higgs mechanism, but I will come back to this later.

Be that as it may, the point I wish to make from this exchange with Pauli is that already in early 1957, just after the first set of parity experiments, many ideas coming to fruition now, had started to become clear. These are:

(1) First was the idea of chiral symmetry leading to a $V-A$ theory. In those early days my humble suggestion (Salam, 1957a, b) of this was limited to neutrinos, electrons, and muons only, while shortly after, that year, Marshak and Sudarshan (Marshak and Sudarshan, 1957 and 1958)[2] Feynman and Gell-Mann (Feynman and Gell-Mann, 1958), and Sakurai (1958) had the courage to postulate γ_5 symmetry for baryons as well as leptons, making this into a universal principle of physics.[3]

Concomitant with the $(V-A)$ theory was the result that if weak interactions are mediated by intermediate mesons, these mesons must carry spin one.

(2) Second was the idea of spontaneous breaking of chiral symmetry to generate electron and muon masses, though the price which those latter-day Shylocks, Nambu and Jona-Lasinio (Nambu and Jona-Lasinio, 1961) and Goldstone [Nambu (1960) and Goldstone (1961)] exacted for this (i.e., the appearance of massless scalars), was not yet appreciated.

(3) And finally, though the use of a Yang–Mills–Shaw (non-Abelian) gauge theory for describing spin-one intermediate charged mesons was suggested already in 1957, the giving of masses to the intermediate bosons through spontaneous symmetry breaking, in such a manner as to preserve the renormalizability of the theory, was to be accomplished only during a long period of theoretical development between 1963 and 1971.

Once the Yang–Mills–Shaw ideas were accepted as relevant to the charged weak currents—to which the charged intermediate mesons were coupled in this theory—during 1957 and 1958 was raised the question of what was the third component of the SU(2) triplet, of which the charged weak currents were the two members. There were the two alternatives: the electroweak unification suggestion, where the electromagnetic current was assumed to be this third component; and the rival suggestion that the third component was a neutral current unconnected with electroweak unification. With hindsight, I shall call these the Klein (1938) (see Klein, 1939) and the Kemmer (1937) alternatives. The Klein suggestion, made in the context of a Kaluza-Klein five-dimensional space-time, was a real tour-de-force; it combined two hypothetical spin-one charged mesons with the photon in one multiplet, deducing from the compactification of the fifth dimension, a theory which looks like Yang–Mills–Shaw's. Klein intended his charged mesons for *strong* interactions, but if we read charged *weak* mesons for Klein's *strong* ones, one obtains the theory independently suggested by Schwinger (1957), though Schwinger, unlike Klein, did not build in any non-Abelian gauge aspects. With just these non-Abelian Yang–Mills gauge aspects very much to the fore, the idea of uniting weak interactions with electromagnetism was developed by Glashow (1959) and Ward and myself (Salam and Ward, 1959) in late 1958. The rival Kemmer suggestion of a global SU(2)-invariant triplet of weak charged and neutral currents was independently suggested by Bludman (1958) in a gauge context and this is how matters stood till 1960.

To give you the flavor of, for example, the year 1960,

[2]The idea of a universal Fermi interaction for (P,N), (ν_e,e), and (ν_μ,μ) doublets goes back to Tiomno and Wheeler (1949a, b) and Yang and Tiomno (1950). Tiomno (1956) considered γ_5 transformations of Fermi fields linked with mass reversal.

[3]Today we believe protons and neutrons are composites of quarks, so that γ_5 symmetry is now postulated for the elementary entities of today—the quarks. If the neutrino also turns out to be massive, γ_5-symmetry is spontaneously broken for it, as it is for electrons, muons, and quarks.

there was a paper written that year by Ward and myself (Salam and Ward, 1961) with the statement "Our basic postulate is that it should be possible to generate strong, weak and electromagnetic interaction terms with all their correct symmetry properties (as well as with clues regarding their relative strengths) by making local gauge transformations on the kinetic energy terms in the free Lagrangian for all particles. This is the statement of an ideal which, in this paper at least, is only very partially realized." I am not laying a claim that we were the only ones who were saying this, but I just wish to convey to you the temper of the physics of twenty years ago—qualitatively no different today from then. But what a quantitative difference the next twenty years made, first with new and far-reaching developments in theory—and then, thanks to CERN, Fermilab, Brookhaven, Argonne, Serpukhov, and SLAC, in testing it!

So far as theory itself is concerned, it was the next seven years between 1961–67 which were the crucial years of quantitative comprehension of the phenomenon of spontaneous symmetry breaking and the emergence of the SU(2)× U(1) theory in a form capable of being tested. The story is well known and Steve Weinberg has already spoken about it. So I will give the barest outline. First there was the realization that the two alternatives mentioned above, a pure electromagnetic current versus a pure neutral current—Klein–Schwinger versus Kemmer–Bludman—were not alternatives; they were complementary. As was noted by Glashow (1961) and independently by Ward and myself (Salam and Ward, 1964), both types of currents and the corresponding gauge particles ($W^{\pm}$, Z^0, and γ) were needed in order to build a theory that could simultaneously accommodate parity violation for weak and parity conservation for the electromagnetic phenomena. Second, there was the influential paper of Goldstone in 1961 which, utilizing a nongauge self-interaction between scalar particles, showed that the price of spontaneous breaking of a continuous internal symmetry was the appearance of zero mass scalars—a result foreshadowed earlier by Nambu. In giving a proof of this theorem (Goldstone *et al.*, 1962) with Goldstone, I collaborated with Steve Weinberg, who spent a year at Imperial College in London. I would like to pay here a most sincerely felt tribute to him and to Sheldon Glashow for their warm and personal friendship.

I shall not dwell on the now well-known contributions of Anderson (1963), Higgs (1964a, 1964b, 1966), Brout and Englert (Englert and Brout, 1964; Englert *et al.*, 1966), Guralnik, Hagen, and Kibble (1964; Kibble, 1967) starting from 1963, which showed how spontaneous symmetry breaking using spin-zero fields could generate vector-meson masses, defeating Goldstone at the same time. This is the so-called Higgs mechanism.

The final steps towards the electroweak theory were taken by Weinberg (1967) and by myself (Salam, 1968) (with Kibble at Imperial College tutoring me about the Higgs phenomena). We were able to complete the present formulation of the spontaneously broken SU(2)× U(1) theory so far as leptonic weak interactions were concerned—with one parameter $\sin^2\theta$ describing all weak and electromagnetic phenomena and with one isodoublet Higgs multiplet. An account of this development was given during the contribution (Salam, 1968) to the Nobel Symposium (organized by Nils Svartholm and chaired by Lamek Hulthén held at Gothenburg after some postponements, in early 1968). As is well known, we did not have then, and still do not have, a prediction for the scalar Higgs mass.

Both Weinberg and I suspected that this theory was likely to be renormalizable.[4] Regarding spontaneously broken Yang–Mills–Shaw theories in general this had earlier been suggested by Englert, Brout, and Thiry (1966). But this subject was not pursued seriously except at Veltman's school at Utrecht, where the proof of renormalizability was given by 't Hooft (1971a, b) in 1971. This was elaborated further by that remarkable physicist, the late Benjamin Lee (Lee, 1972; Lee and Zinn-Justin, 1972, 1973), working with Zinn-Justin, and by 't Hooft and Veltman (1972a, 1972b).[5] This followed on the earlier basic advances in Yang–Mills calculational technology by Feynman (1963), DeWitt (1967a, b), Faddeev and Popov (1967), Mandelstam (1968a, b), Fradkin and Tyutin (1970), Boulware (1970), Taylor (1971), Slavnov (1972), Strathdee and Salam (Salam and Strathdee, 1970). In Coleman's eloquent phrase "'t Hooft's work turned the Weinberg–Salam frog into an enchanted prince." Just before had come the GIM (Glashow, Iliopoulos, and Maiani) mechanism (Glashow *et al.*, 1970), emphasizing that the existence of the fourth charmed quark (postulated earlier by several authors) was essential to the natural resolution of the dilemma posed by the absence of strangeness-violating currents. This tied in naturally with the understanding of the Steinberger–Schwinger–Rosenberg–Bell–Jackiw–Adler anomaly (see Jackiw, 1972) and its removal for SU(2) × U(1) by the parallelism of four quarks and four leptons, pointed out by Bouchiat, Iliopoulos, and Meyer (1972) and independently by Gross and Jackiw (1972).

If one has kept a count, I have so far mentioned around fifty theoreticians. As a failed experimenter, I have always felt envious of the ambience of large experimental teams and it gives me the greatest pleasure to acknowledge the direct or the indirect contributions of the "series of minds" to the spontaneously broken SU(2)× U(1) gauge theory. My profoundest personal appreciation goes to my collaborators at Imperial College, and Cambridge and the Trieste Centre, John Ward, Paul Matthews, Jogesh Pati, John Strathdee, Tom Kibble, and to Nicholas Kemmer.

In retrospect, what strikes me most about the early part of this story is how uninformed all of us were, not

[4]When I was discussing the final version of the SU(2)×U(1) theory and its possible renormalizability in Autumn 1967 during a postdoctoral course of lectures at Imperial College, Nino Zichichi from CERN happened to be present. I was delighted because Zichichi had been badgering me since 1958 with persistent questioning as to of what theoretical avail his precise measurements on (g–2) for the muon as well as those of the muon lifetime were, when not only the magnitude of the electromagnetic corrections to weak decays was uncertain, but also conversely the effect of nonrenormalizable weak interactions on "renormalized" electromagnetism was so unclear.

[5]An important development in this context was the invention of the dimensional regularization technique by Bollini and Giambiagi (1972), Ashmore (1972), and 't Hooft and Veltman.

only of each other's work, but also of work done earlier. For example, only in 1972 did I learn of Kemmer's paper written at Imperial College in 1937. Kemmer's argument essentially was that Fermi's weak theory was not globally SU(2) invariant and should be made so—though not for its own sake but as a prototype for strong interactions. Then this year I learnt that earlier, in 1936, Kemmer's Ph.D. supervisor, Gregor Wentzel (1937), had introduced (the yet undiscovered) analogs of lepto-quarks, whose mediation could give rise to neutral currents after a Fierz reshuffle. And only this summer, Cecilia Jarlskog at Bergen rescued Oscar Klein's paper from the anonymity of the Proceedings of the International Institute of Intellectual Cooperation of Paris, and we learnt of his anticipation of a theory similar to Yang–Mills–Shaw long before these authors. As I indicated before, the interesting point is that Klein was using his triplet, of two charged mesons plus the photon, not to describe weak interaction but for strong nuclear force unification with the electromagnetic—something our generation started on only in 1972—and not yet experimentally verified. Even in this recitation I am sure I have inadvertently left off some names of those who have in some way contributed to SU(2)× U(1). Perhaps the moral is that not unless there is the prospect of quantitative verification, does a qualitative idea make its impress in physics.

And this brings me to experiment, and the year of the Gargamelle (Hasert *et al.*, 1973). I still remember Paul Matthews and I getting off the train at Aix-en-Provence for the 1973 European Conference and foolishly deciding to walk with our rather heavy luggage to the student hostel where we were billeted. A car drove from behind us, stopped, and the driver leaned out. This was Musset whom I did not know well personally then. He peered out of the window and said: "Are you Salam?" I said "Yes." He said: "Get into the car. I have news for you. We have found neutral currents." I will not say whether I was more relieved for being given a lift because of our heavy luggage or for the discovery of neutral currents. At the Aix-en-Provence meeting that great and modest man, Lagarrigue, was also present and the atmosphere was that of a carnival—at least this is how it appeared to me. Steve Weinberg gave the rapporteur's talk with T. D. Lee as the chairman. T. D. was kind enough to ask me to comment after Weinberg finished. That summer Jogesh Pati and I had predicted proton decay within the context of what is now called grand unification, and in the flush of this excitement I am afraid I ignored weak neutral currents as a subject which had already come to a successful conclusion, and concentrated on speaking of the possible decays of the proton. I understand now that proton decay experiments are being planned in the United States by the Brookhaven, Irvine and Michigan and the Wisconsin–Harvard groups and also by a European collaboration to be mounted in the Mont Blanc Tunnel Garage No. 17. The later quantitative work on neutral currents at CERN, Fermilab, Brookhaven, Argonne and Serpukhov is, of course, history, but a special tribute is warranted to the beautiful SLAC–Yale–CERN experiment (Taylor, 1979) of 1978 which exhibited the effective Z^0-photon interference in accordance with the predictions of the theory. This was foreshadowed by Barkov *et al.*'s experiments (Barkov, 1979) at Novosibirsk in the USSR in their exploration of parity-violation in the atomic potential for bismuth. There is the apocryphal story about Einstein, who was asked what he would have thought if experiment had not confirmed the light deflection predicted by him. Einstein is supposed to have said, "Madam, I would have thought the Lord has missed a most marvelous opportunity." I believe, however, that the following quote from Einstein's Herbert Spencer lecture of 1933 expresses his, my colleagues', and my own views more accurately. "Pure logical thinking cannot yield us any knowledge of the empirical world; all knowledge of reality starts from experience and ends in it." This is exactly how I feel about the Gargamelle–SLAC experience.

III. THE PRESENT AND ITS PROBLEMS

Thus far we have reviewed the last twenty years and the emergence of SU(2)× U(1), with the twin developments of a gauge theory of basic interactions, linked with internal symmetries, and of the spontaneous breaking of these symmetries. I shall first summarize the situation as we believe it to exist now and the immediate problems. Then we turn to the future.

(1) To the level of energies explored, we believe that the following sets of particles are "structureless" (in a field-theoretic sense) and, at least to the level of energies explored hitherto, constitute the elementary entities of which all other objects are made.

$$\mathrm{SU_c(3)\ triplets}$$

$$\text{Family I}\quad \text{quarks}\begin{Bmatrix} u_R, u_Y, u_B \\ d_R, d_Y, d_B \end{Bmatrix}\quad \text{leptons}\begin{bmatrix} \nu_e \\ e \end{bmatrix}\quad \text{SU(2)doublets}$$

$$\text{Family II}\quad \text{quarks}\begin{Bmatrix} c_R, c_Y, c_B \\ s_R, s_Y, s_B \end{Bmatrix}\quad \text{leptons}\begin{bmatrix} \nu_\mu \\ \mu \end{bmatrix}\quad \text{SU(2)doublets}$$

$$\text{Family III}\quad \text{quarks}\begin{Bmatrix} t_R, t_Y, t_B \\ b_R, b_Y, b_B \end{Bmatrix}\quad \text{leptons}\begin{bmatrix} \nu_\tau \\ \tau \end{bmatrix}\quad \text{SU(2)doublets}$$

Together with their antiparticles each family consists of 15 or 16 two-component fermions (15 or 16 depending on whether the neutrino is four-component or not). The third family is still conjectural, since the top quark (t_R, t_Y, t_B) has not yet been discovered. Does this family really follow the pattern of the other two? Are there more families? Does the fact that the families are replicas of each other imply that Nature has discovered a dynamical stability about a system of 15 (or 16) objects, and that by this token there is a more basic layer of structure underneath? (See Pati and Salam, 1975a; Pati *et al.*, 1975a; Harari, 1979; Schupe, 1979; Curtwright and Fruend, 1979).

(2) Note that quarks come in three colors: Red (R), Yellow (Y), and Blue (B). Parallel with the electroweak SU(2)× U(1), a gauge field[6] theory ($\mathrm{SU_c(3)}$) of

strong (quark) interactions (quantum chromodynamics, QCD[7] has emerged which gauges the three colors. The indirect discovery of the (eight) gauge bosons associated with QCD (gluons), has already been surmised by the groups at DESY.[8]

(3) All known baryons and mesons are singlets of color $SU_c(3)$. This has led to a hypothesis that color is always confined. One of the major unsolved problems of field theory is to determine if QCD—treated non-perturbatively—is capable of confining quarks and gluons.

(4) In respect of the electroweak $SU(2)\times U(1)$, all known experiments on weak and electromagnetic phenomena below 100 GeV carried out to date agree with the theory which contains one theoretically undetermined parameter $\sin^2\theta = 0.230 \pm 0.009$ (Winter, 1979). The predicted values of the associated gauge boson ($W^{\pm}$ and Z^0) masses are: $m_W \approx 77$–84 GeV, $m_Z \approx 89$–95 GeV, for $0.25 \geq \sin^2\theta \geq 0.21$.

(5) Perhaps the most remarkable measurement in electroweak physics is that of the parameter $\rho = (m_W/m_Z \cos\theta)^2$. Currently this has been determined from the ratio of neutral to charged current cross sections. The predicted value $\rho = 1$ for weak *iso-doublet Higgs* is to be compared with the experimental[9] $\rho = 1.00 \pm 0.02$.

(6) Why does Nature favor the simplest suggestion in $SU(2)\times U(1)$ theory of the Higgs scalars being iso-doublet?[10] Is there just one physical Higgs? Of what mass? At present the Higgs interactions with leptons and quarks as well as their self-interactions are non-gauge interactions. For a three-family (six-quark) model, 21 out of the 26 parameters needed are attributable to the Higgs interactions. Is there a basic principle, as compelling and as economical as the gauge principle, which embraces the Higgs sector? Alternatively, could the Higgs phenomenon itself be a manifestation of a dynamical breakdown of the gauge symmetry?[10]

(7) Finally there is the problem of the families; is there a distinct SU(2) for the first, another for the second, as well as a third SU(2), with spontaneous symmetry breaking such that the SU(2) apprehended by present experiment is a diagonal sum of these "family" SU(2)'s? To state this in another way, how far in energy does the $e-\mu$ universality (for example) extend? Are there more[11] Z^0's than just one, effectively differentially coupled to the e and the μ systems? (If there are, this will constitute minimodifications of the theory, but not a drastic revolution of its basic ideas.)

In the next section I turn to a direct extrapolation of the ideas which went into the electroweak unification, so as to include strong interactions as well. Later I shall consider the more drastic alternatives which may be needed for the unification of all forces (including gravity)—ideas which have the promise of providing a deeper understanding of the charge concept. Regretfully, by the same token, I must also become more technical and obscure for the nonspecialist. I apologize for this. The nonspecialist may sample the flavor of the arguments in the next section (Sec. IV), ignore the Appendices, and then go on to Sec. V, which is perhaps less technical.

IV. DIRECT EXTRAPOLATION FROM THE ELECTROWEAK TO THE ELECTRONUCLEAR

A. The three ideas

The three main ideas which have gone into the electronuclear—also called grand—unification of the electroweak with the *strong* nuclear force (and which date back to the period 1972–1974), are the following:

(1) First: the psychological break (for us) of grouping quarks and leptons in the *same* multiplet of a unifying group G, suggested by Pati and myself in 1972 (see Bjorken, 1972; Pati and Salam, 1973a). The group G must contain $SU(2)\times U(1)\times SU_c(3)$, and must be non-Abelian, if all quantum numbers (flavor, color, lepton, quark, and family numbers) are to be automatically quantized and the resulting gauge theory asymptotically free.

(2) Second: an extension, proposed by Georgi and Glashow (1974) which places not only (left-handed) quarks and leptons but also their antiparticles in the same multiplet of the unifying group.

Appendix I displays some examples of the unifying groups presently considered.

[6]"To my mind the most striking feature of theoretical physics in the last thirty-six years is the fact that not a single new theoretical idea of a fundamental nature has been successful. The notions of relativistic quantum theory.... have in every instance proved stronger than the revolutionary ideas....of a great number of talented physicists. We live in a dilapidated house and we seem to be unable to move out. The difference between this house and a prison is hardly noticeable"—Res Jost (1963), "In Praise of Quantum Field Theory" (Siena European Conference).

[7]Pati and Salam. See the review by Bjorken (1972). See also Fritzsch and Gell-Mann (1972), Fritzsch, Gell-Mann, and Leutwyler (1973), Weinberg (1973a,b), and Gross and Wilczek (1973). For a review see Marciano and Pagels (1978).

[8]See the Tasso Collaboration (Brandelik *et al.*, 1979) and the Mark-J Collaboration (Barber *et al.*, 1979). See also the reports of the Jade, Mark-J, Pluto, and Tasso Collaborations to the International Symposium on Lepton and Photon Interactions at High Energies, Fermilab, August 1979.

[9]The one-loop radiative corrections to ρ suggest that the maximum mass of leptons contributing to ρ is less than 100 GeV (Ellis, 1979).

[10]To reduce the arbitrariness of the Higgs couplings and to motivate their iso-doublet character, one suggestion is to use supersymmetry. Supersymmetry is a Fermi-Bose symmetry, so that iso-doublet leptons like (ν_e, e) or (ν_μ, μ) in a supersymmetric theory must be accompanied in the same multiplet by iso-doublet Higgs. Alternatively, one may identify the Higgs as composite fields associated with bound states of a yet new level of elementary particles and new (so-called techni-color) forces (Dimopoulos and Susskind, 1979; Weinberg, 1979a; and 't Hooft) of which, at present low energy, we have no cognizance, and which may manifest themselves in the 1–100 TeV range. Unfortunately, both these ideas at first sight appear to introduce complexities, though in the context of a wider theory, which spans energy scales up to much higher masses, a satisfactory theory of the Higgs phenomena, incorporating these, may well emerge.

[11]See Pati and Salam (1974); Mohapatra and Pati (1975a,b); Elias, Pati, and Salam (1978a); and Pati and Rajpoot (1978).

Now a gauge theory based on a "simple" (or with discrete symmetries, a "semisimple") group G contains one basic gauge constant. This constant would manifest itself physically above the "grand unification mass" M, exceeding all particle masses in the theory—these themselves being generated (if possible) hierarchially through a suitable spontaneous symmetry-breaking mechanism.

(3) The third crucial development was by Georgi, Quinn, and Weinberg (1974) who showed how, using renormalization group ideas, one could relate the observed low-energy couplings $\alpha(\mu)$ and $\alpha_s(\mu)(\mu \sim 100\,\text{GeV})$ to the magnitude of the grand unifying mass M and the observed value of $\sin^2\theta(\mu)$; ($\tan\theta$ is the ratio of the U(1) to the SU(2) couplings).

(4) If one extrapolates with Jowett,[12] that nothing essentially new can possibly be discovered—i.e., if one assumes that there are no new features, no new forces, or no new "types" of particles to be discovered, till we go beyond the grand unifying energy M—then the Georgi, Quinn, Weinberg method leads to a startling result: this featureless "plateau" with no "new physics" heights to be scaled stretches to fantastically high energies. More precisely, if $\sin^2\theta(\mu)$ is as large as 0.23, then the grand unifying mass M cannot be smaller than 1.3×10^{13} GeV (Marciano, 1979). (Compare with Planck mass $m_P \approx 1.2\times10^{19}$ GeV related to Newton's constant where gravity must come in.)[13] The result follows from the formula (Marciano, 1979; Salam, 1979).

$$\frac{11\alpha}{3\pi}\ln\frac{M}{\mu} = \frac{\sin^2\theta(M) - \sin^2\theta(\mu)}{\cos^2\theta(M)}, \tag{1}$$

if it is assumed that $\sin^2\theta(M)$—the magnitude of $\sin^2\theta$ for energies of the order of the unifying mass M—equals 3/8 (see Appendix B).

This startling result will be examined more closely in Appendix B. I show there that it is very much a consequence of the assumption that the $SU(2)\times U(1)$ symmetry survives intact from the low regime energies μ right upto the grand unifying mass M. I will also show that there already is some experimental indication that this assumption is too strong, and that there may be likely peaks of new physics at energies of 10 TeV upwards.

B. Tests of electronuclear grand unification

The most characteristic prediction from the existence of the electronuclear force is proton decay, first discussed in the context of grand unification at the Aix-en-Provence Conference (1973) (Pati and Salam, 1973b). For "semisimple" unifying groups with multiplets containing quarks and leptons only (but no antiquarks nor antileptons) the lepto-quark composites have masses (determined by renormalization group arguments) of the order of $\approx 10^5$–10^6 GeV (Elias *et al.*, 1978b; Rajpoot and Elias, 1978). For such theories the characteristic proton decays (proceeding through exchanges of *three* lepto-quarks) conserve quark number + lepton number, i.e., $P = qqq \to lll$, $\tau_P \sim 10^{29}$–10^{34} years. On the contrary, for the "simple" unifying family groups like SU(5) (Georgi and Glashow, 1974) or SO(10) (Fritzsch and Minkowski, 1975, 1976; Georgi, 1975; Georgi and Nanopoulos, 1979) (with multiplets containing antiquarks and antileptons) proton decay proceeds through an exchange of *one* lepto-quark into an antilepton (plus pions, etc.) ($P \to \bar{l}$).

An intriguing possibility in this context is that investigated recently for the maximal unifying group SU(16)—the largest group to contain a sixteenfold fermionic family $(q, l, \bar{q}, \bar{l})$. This can permit four types of decay modes: $P \to 3l$ *as well as* $P \to \bar{l}$, $P \to l$ (e.g., $P \to l^- + \pi^+ + \pi^+$), and $P \to 3\bar{l}$ (e.g., $N \to 3\bar{\nu} + \pi^0$, $P \to 2\bar{\nu} + e^+ + \pi^0$), the relative magnitudes of these alternative decays being model-dependent on how precisely SU(16) breaks down to $SU(3)\times SU(2)\times U(1)$. Quite clearly, it is the central fact of the existence of the proton decay for which the present generation of experiments must be designed, rather than for any specific type of decay modes.

Finally, grand unifying theories predict mass relations like (Buras *et al.*, 1978):

$$\frac{m_d}{m_e} = \frac{m_s}{m_\mu} = \frac{m_b}{m_\tau} \approx 2.8$$

for six (or at most eight) flavors *below the unification mass*. The important remark for proton decay and for mass relations of the above type as well as for an understanding of baryon excess[14] in the universe,[15] is that for the present *these are essentially characteristic of the fact of grand unification—rather than of specific models*.

"Yet each man kills the thing he loves" sang Oscar Wilde anguishedly in his famous Ballad of the Reading

---

[12] The universal urge to extrapolate from what we know today and to believe that nothing new can possibly be discovered is well expressed in the following:

"I come first, My name is Jowett
I am the Master of this College,
Everything that is, I know it
If I don't, it isn't knowledge"

— The Balliol Masque.

[13] On account of the relative proximity of $M \approx 10^{13}$ GeV to m_P (and the hope of eventual unification with gravity), Planck mass m_P is now the accepted "natural" mass scale in particle physics. With this large mass as the input, the great unsolved problem of grand unification is the "natural" emergence of mass hierarchies (m_P, αm_P, $\alpha^2 m_P$, ...) or $m_P \exp(-c_n/\alpha)$, where c_n's are constants. [$m_e/m_P \sim 10^{-22}$.]

[14] See Yoshimura (1978), Dimopoulos and Susskind (1978), Toussiant *et al.* (1979), Ellis *et al.* (1979), Weinberg (1979b), and Nanopoulos and Weinberg (1979).

[15] The calculation of baryon excess in the universe—arising from a combination of CP and baryon number violations—has recently been claimed to provide teleological arguments for grand unification. For example, Nanopoulos (1979) has suggested that the "existence of human beings to measure the ratio n_B/n_γ (where n_B is the number of baryons and n_γ the number of photons in the universe) necessarily imposes severe bounds on this quantity: i.e., $10^{-11} \approx (m_e/m_P)^{1/2} \lesssim n_B/n_\gamma \lesssim 10^{-4}$ ($\approx 0(\alpha^2)$)." Of importance in deriving these constraints are the upper (and lower) bounds on the numbers of flavors (≈ 6) deduced (1) from mass relations above, (2) from cosmological arguments which seek to limit the numbers of massless neutrinos, (3) from asymptotic freedom, and (4) from numerous (one-loop) radiative calculations. It is clear that lack of accelerators as we move up in energy scale will force particle physics to reliance on teleology and cosmology (which, in Landau's famous phrase, is "often wrong, but never in doubt").

Gaol. Like generations of physicists before us, some in our generation also (through a direct extrapolation of the electroweak gauge methodology to the electronuclear)—and with faith in the assumption of no "new physics," which leads to a grand unifying mass $\sim 10^{13}$ GeV—are beginning to believe that the end of the problems of elementarity as well as of fundamental forces is nigh. They may be right, but before we are carried away by this prospect, it is worth stressing that even for the simplest grand unifying model [Georgi and Glashow's SU(5) with just two Higgs (a 5 and a 24)], the number of presently *ad hoc* parameters needed by the model is still unwholesomely large—22, compared with 26 for the six-quark model based on the humble $SU(2)\times U(1)\times SU_c(3)$. We cannot feel proud.

V. ELEMENTARITY: UNIFICATION WITH GRAVITY AND NATURE OF CHARGE

In some of the remaining parts of this lecture I shall be questioning two of the notions which have gone into the direct extrapolation of Sec. IV—first, do quarks and leptons represent the correct elementary[16] fields, which should appear in the matter Lagrangian, and which are structureless for renormalizability; second, could some of the presently considered gauge fields themselves be composite? This part of the lecture relies heavily on an address I was privileged to give at the European Physical Society meeting in Geneva in July this year (Salam, 1979).

A. The quest for elementarity, prequarks (preons and pre-preons)

If quarks and leptons are elementary, we are dealing with $3\times 15 = 45$ elementary entities. The "natural" group of which these constitute the fundamental representation is SU(45) with 2024 elementary gauge bosons. It is possible to reduce the size of this group to SU(11) for example (see Appendix A) with only 120 gauge bosons, but then the number of fermions increases to 501 (of which presumably $3\times 15 = 45$ objects are of low and the rest of Planckian mass). Is there any basic reason for one's instinctive revulsion when faced with these vast numbers of elementary fields?

The numbers by themselves would perhaps not matter so much. After all, Einstein, in his description of gravity (Einstein, 1916), chose to work with 10 fields $[g_{\mu\nu}(x)]$ rather than with just one (scalar field) as Nördstrom [(1912; 1913a, b; 1914a, b); see also Einstein (1912a, b)] before him. Einstein was not perturbed by the multiplicity he chose to introduce, since he relied on the sheet-anchor of a fundamental principle (the equivalence principle) which permitted him to relate the ten fields for gravity $g_{\mu\nu}$ with the ten components of the physically relevant quantity, the tensor $T_{\mu\nu}$ of energy and momentum. *Einstein knew that nature was not economical of structures:* only of principles of fundamental applicability. The question we must ask ourselves is this: Have we yet discovered such principles in our quest for elementarity, to justify having fields with such large numbers of components as elementary?

Recall that quarks carry at least three charges (color, flavor, and a family number). Should one not, by now, entertain the notions of quarks (and possibly of leptons) as being composites of some more basic entities[17] (prequarks or preons), which each carry but *one* basic charge? (Pati and Salam, 1975a; Pati *et al.*, 1975a; Harari, 1979; Schupe, 1979; Curtright and Freund, 1979) These ideas have been expressed before but they have become more compulsive now, with the growing multiplicity of quarks and leptons. Recall that it was similar ideas which led from the eightfold of baryons to a triplet of (Sakatons and) quarks in the first place.

The preon notion is now new. In 1975, among others, Pati, Salam, and Strathdee (1975a) introduced 4 chromons (the fourth color corresponding to the lepton number) and 4 flavons, the basic group being SU(8)—of which the family group $SU_F(4)\times SU_C(4)$ was but a subgroup. As an extension of these ideas, we now believe these preons carry magnetic charges and are bound together by very strong short-range forces, with quarks and leptons as their magnetically neutral composites (Pati and Salam, 1980). The important remark in this context is that in a theory containing *both* electric and magnetic generalized charges, the analogs of the well-known Dirac quantization condition (Dirac, 1931) give relations like $eg/4\pi = n/2$ for the strength of the two types of charges. Clearly, magnetic monopoles[18] of strength g and mass $mw/\alpha \approx 10^4 - 10^5$ GeV of opposite polarity, are likely to bind much more tightly than electric charges, yielding composites whose nonelementary nature will reveal itself only for very high energies. This appears to be the situation at least for leptons if they are composites.

In another form the preon idea has been revived this year by Curtright and Freund (1979), who, motivated by ideas of extended supergravity (to be discussed in the next subsection), reintroduce an SU(8) of 3 chromons (R, Y, B), 2 flavons, and 3 familons (horrible names). The family group SU(5) could be a subgroup of this SU(8). In the Curtright–Freund scheme, the $3\times 15 = 45$ fermions of SU(5) (Georgi and Glashow, 1974) can be found among the $8 + 28 + 56$ of SU(8) [or alternatively the $3\times 16 = 48$ of SO(10) among the vectorial 56 fermions of SU(8)]. (The next succession after the preon level may be the pre-preon level. It was suggested at the Geneva

[16] I would like to quote Feynman in a recent interview in *Omni* magazine: "As long as it looks like the way things are built [is] with wheels within wheels, then you are looking for the innermost wheel—but it might not be that way, in which case you are looking for whatever the hell it is you find!" In the same interview he remarks "a few years ago I was very skeptical about the gauge theories.... I was expecting mist, and now it looks like ridges and valleys after all."

[17] One must emphasize, however, that zero mass neutrinos are the hardest objects to conceive of as composites.

[18] According to 't Hooft's theorem, a monopole corresponding to the $SU_L(2)$ gauge symmetry is expected to possess a mass with the lower limit m_W/α ('t Hooft, 1974; Polyakov, 1974). Even if such monopoles are confined, their indirect effects must manifest themselves, if they exist. (Note that m_W/α is very much a lower limit for a grand unified theory like SU(5), for which the monopole mass is α^{-1} times the heavy lepto-quark mass.) The monopole force may be the techni-force of Footnote 10.

TABLE I. Prognosis for the next decade.

Decade	1950–1960	1960–1970	1970–1980	1980→
Discovery in early part of the decade	The strange particles	The 8-fold way, Ω^-	Confirmation of neutral currents	W, Z, Proton decay
Expectation for the rest of the decade		SU(3) resonances		Grand Unification, Tribal Groups
Actual discovery		Hit the next level of elementarity with quarks		May hit the preon level, and composite structure of quarks

Conference (see Salam, 1979) that with certain developments in field theory of composite fields it could be that just two pre-preons may suffice. But at this stage this is pure speculation.)

Before I conclude this section, I would like to make a prediction regarding the course of physics in the next decade, extrapolating from our past experience of the decades gone by (See Table I).

B. Post-Planck physics, supergravity, and Einstein's dreams

I now turn to the problem of a deeper comprehension of the charge concept (the basis of gauging)—*which, in my humble view, is the real quest of particle physics.* Einstein, in the last 35 years of his life, lived with two dreams: one was to unite gravity with matter (the photon)—he wished to see the "base wood" (as he put it) which makes up the stress tensor $T_{\mu\nu}$ on the right-hand side of his equation $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -T_{\mu\nu}$ transmuted through this union, into the "marble" of gravity on the left-hand side. The second (and the complementary) dream was to use this unification to comprehend the nature of electric charge in terms of space-time geometry in the same manner as he had successfully comprehended the nature of gravitational charge in terms of space-time curvature.

In case someone imagines[19] that such deeper comprehension is irrelevant to quantitative physics, let me adduce the tests of Einstein's theory versus the proposed modifications to it [Brans–Dicke (Brans and Dicke, 1961) for example]. Recently (1976), the *strong* equivalence principle (i.e., the proposition that gravitational forces contribute equally to the inertial and the gravitational masses) was tested[20] to one part in 10^{12} [i.e., to the same accuracy as achieved in particle physics for $(g-2)_e$] through lunar-laser ranging measurements [Williams *et al.* (1976), Shapiro *et al.* (1976). For a discussion see Salam (1977)]. These measurements determined departures from Kepler equilibrium distances of the moon, the earth, and the sun to better than ±30 cm and triumphantly vindicated Einstein.

There have been four major developments in realizing Einstein's dreams:

(1) The Kaluza-Klein (Kaluza, 1921; Klein, 1926) miracle: An Einstein Lagrangian (scalar curvature) in five-dimensional space-time (where the fifth dimension is compactified in the sense of all fields being explicitly independent of the fifth coordinate) precisely reproduces the *Einstein-Maxwell* theory in four dimensions, the $g_{\mu 5}(\mu = 0, 1, 2, 3)$ components of the metric in five dimensions being identified with the Maxwell field A_μ. From this point of view, Maxwell's field is associated with the extra components of curvature implied by the (conceptual) existence of the fifth dimension.

(2) The second development is the recent realization by Cremmer, Scherk, Englert, Brout, Minkowski, and others that the compactification of the extra dimensions (Cremmer and Scherk, 1976a, b, c; Minkowski, 1977)—(their curling up to sizes perhaps smaller than Planck length $\lesssim 10^{-33}$ cm and the very high curvature associated with them)—might arise through a spontaneous symmetry breaking (in the first 10^{-43} sec) which reduced the higher-dimensional space-time effectively to the four-dimensional that we apprehend directly.

(3) So far we have considered Einstein's second dream, i.e., the unification of electromagnetism (and presumably of other gauge forces) with gravity, giving a space-time significance to gauge charges as corresponding to extended curvature in extra bosonic dimensions. A full realization of the first dream (unification of spinor matter with gravity and with other gauge fields) had to await the development of supergravity[21]—and an extension to extra fermionic dimensions of superspace (Salam and Strathdee, 1974a) (with extended torsion being brought into play in addition to curvature). I discuss this development later.

(4) And finally there was the alternative suggestion by Wheeler (Fuller and Wheeler, 1962; Wheeler, 1964) and Schemberg that electric charge may be associated with space-time topology—with worm-holes, with space-time Gruyère-cheesiness. This idea has re-

[19]The following quotation from Einstein is relevant here. "We now realize, with special clarity, how much in error are those theorists who believe theory comes inductively from experience. Even the great Newton could not free himself from this error (*Hypotheses non fingo*)." This quote is complementary to the quotation from Einstein at the end of Sec. II.

[20]The *weak* equivalence principle (the proposition that all but the gravitational force contribute equally to the inertial and the gravitational masses) was verified by Eötvös to $1:10^8$ and by Dicke and Braginsky and Panov to $1:10^{12}$.

[21]See Freedman, van Nieuwenhuizen, and Ferrara (1976) and Deser and Zumino (1976). For a review and comprehensive list of references, see Freedman (1979). See also Arnowitt, Nath, and Zumino (1975), Zumino (1975), Wess and Zumino (1977), Akulov, Volkov, and Soroka (1975), and Brink *et al.* (1978).

cently been developed by Hawking[22] and his collaborators (Hawking, 1978, 1979a, b; Gibbons *et al.*, 1978).

C. Extended supergravity, SU(8) preons, and composite gauge fields

Now so far I have reviewed the developments in respect of Einstein's dreams as reported at the Stockholm Conference held in 1978 in this hall and organized by the Swedish Academy of Sciences.

A remarkable new development was reported during 1979 by Julia and Cremmer (Cremmer *et al.* (1978); Cremmer and Julia (1978, 1979); see also Julia (1979)] which started with an attempt to use the ideas of Kaluza and Klein to formulate extended supergravity theory in a higher (compactified) space-time—more precisely, in eleven dimensions. This development links up, as we shall see, with preons and composite Fermi fields—and even more important—possibly with the notion of composite gauge fields.

Recall that simple supergravity[23] is the gauge theory of supersymmetry[24]—the gauge particles being the (helicity ±2) gravitons and (helicity $\pm\frac{3}{2}$) gravitinos.[25] *Extended supergravity* gauges supersymmetry combined with SO(N) internal symmetry. For $N=8$, the (tribal) supergravity multiplet consists of the following SO(8) families.[26]

Helicity	
±2	1
$\pm\frac{3}{2}$	8
±1	28
$\pm\frac{1}{2}$	56
0	70

As is well known, SO(8) is too small to contain SU(2) × U(1) × $SU_c(3)$. Thus this tribe has no place for $W^{\pm}$ (though Z^0 and γ are contained) and no places for μ or τ or the t quark.

This was the situation last year. This year, Cremmer and Julia (see Footnote 26) attempted to write down the $N=8$ supergravity Lagrangian explicitly, using an extension of the Kaluza–Klein ansatz which states that *extended supergravity* [with SO(8) internal symmetry] has the same Lagrangian in four space-time dimensions as *simple supergravity* in (compactified) eleven dimensions. This formal—and rather formidable—ansatz, when carried through, yielded a most agreeable bonus. *The supergravity Lagrangian possesses an unsuspected SU(8) "local" internal symmetry* although one started with an internal SO(8) only.

The tantalizing questions which now arise are the following.

(1) Could this internal SU(8) be the symmetry group of the 8 preons (3 chromons, 2 flavons, 3 familons) introduced earlier?

(2) When SU(8) is gauged, there should be 63 spin-one fields. The supergravity tribe contains only 28 spin-one fundamental objects which are not minimally coupled. Are the 63 fields of SU(8) to be identified with composite gauge fields made up of the 70 spin-zero objects of the form $V^{-1}\partial_\mu V$? Do these composites propagate, in analogy with the well-known recent result in CP^{n-1} theories (D'Adda *et al.*, 1978), where a composite gauge field of this form propagates as a consequence of quantum effects (quantum completion)?

The entire development I have described—the unsuspected extension of SO(8) to SU(8) when extra compactified space-time dimensions are used, and the possible existence and quantum propagation of composite gauge fields—is of such crucial importance for the future prospects of gauge theories that one begins to wonder how much of the extrapolation which took SU(2) × U(1) × $SU_c(3)$ into the electronuclear grand unified theories is likely to remain unaffected by these new ideas now unfolding.

But where in all this is the possibility to appeal directly to experiment? For grand unified theories, it was the proton decay. What is the analog for supergravity? Perhaps the spin-$\frac{3}{2}$ massive gravitino, picking its mass from a super-Higgs effect [Cremmer *et al.* (1979); see also Ferrara (1979) and references therein] provides the answer. Fayet (1977, 1979) has shown that for a spontaneously broken globally supersymmetric weak theory the introduction of a local gravitational interaction leads to a super-Higgs effect. Assuming that supersymmetry breakdown is at mass scale m_w, the gravitino acquires a mass and an effective interaction, but of conventional weak rather than of the gravitational strength—an enhancement by a factor of 10^{34}. One may thus search for the gravitino among the neutral decay modes of J/ψ—the predicted rate being 10^{-3}–10^{-5} times smaller than the observed rate for $J/\psi \to e^+e^-$. This will surely tax all the ingenuity of the great men (and women) at SLAC and DESY. Another effect suggested by Scherk (1979) is antigravity—a cancellation of the attractive gravitational force with the force produced by spin-one gravi-photons which exist in all extended supergravity theories. Scherk shows that the Compton wavelength of the gravi-photon is either smaller than 5 cm or is between 10 and 850 meters in order that there will be no conflict with what

---

[22]The Einstein Lagrangian allows large fluctuations of metric and topology on Planck-length scale. Hawking has surmised that the dominant contributions to the path integral of quantum gravity come from metrics which carry one unit of topology per Planck volume. On account of the intimate connection (de Rham, Atiyah-Singer) (Atiyah and Singer, 1963) of curvature with the measures of space-time topology (Euler number, Pontryagin number) the extended Kaluza-Klein and Wheeler-Hawking points of view may find consonance after all.

[23]See Freedman, van Nieuwenhuizen, and Ferrara (1976), and Deser and Zumino (1976). For a review and comprehensive list of references, see Freedman (1979).

[24]See Gol'fand and Likhtman (1971), Volkov and Akulov (1972), Wess and Zumino (1974), Salam and Strathdee (1974a,b,c). For a review, see Salam and Strathdee (1978).

[25]Supersymmetry algebra extends Poincarè group algebra by adjoining to it supersymmetric charges Q_α which transform bosons to fermions. $\{Q_\alpha, Q_\beta\} = (\gamma_\mu P_\mu)_{\alpha\beta}$. The currents which correspond to these charges (Q_α and P_μ) are $J_{\mu\alpha}$ and $T_{\mu\nu}$—these are essentially the currents which in gauged supersymmetry (i.e., supergravity) couple to the gravitino and the graviton, respectively.

[26]See Footnote 23 and Cremmer, Julia, and Scherk (1978) and Cremmer and Julia (1978, 1979). See also Julia (1979).

TABLE A.I. Examples of grand unifying groups.

Semisimple groups[27]	Multiplet	Exotic gauge particles	Proton decay
(with left-right symmetry)	$G_L \to \begin{pmatrix} q \\ l \end{pmatrix}_L$, $G_R \to \begin{pmatrix} q \\ l \end{pmatrix}_R$	Lepto-quarks $\to (\bar{q}l)$	Lepto-quarks $\to W$ + (Higgs) or
Example $[SU(6)_F \times SU(6)_c]_{L \to R}$	$G = G_L \times G_R$	Unifying mass $\approx 10^6$ GeV	Proton $= qqq \to lll$
Examples Family groups $\to$ SU(5) or SO(10) Tribal groups $\to$ SU(11) SO(22)	$G \to \begin{pmatrix} q \\ l \\ \bar{q} \\ \bar{l} \end{pmatrix}_L$	Diquarks $\to (qq)$ Dileptons $\to (ll)$ Lepto-quarks $\to (\bar{q}l), (ql)$ Unifying mass $\approx 10^{13} - 10^{15}$ GeV	$qq \to \bar{q}\bar{l}$ i.e. Proton $P = qqq \to \bar{l}$ Also possible, $P \to l$, $P \to 3\bar{\nu}$, $P \to 3\nu$

is presently known about the strength of the gravitational force.

Let me summarize: it is conceivable of course, that there is indeed a grand plateau—extending even to Planck energies. If so, the only eventual laboratory for particle physics will be the early universe, where we shall have to seek for the answers to the questions on the nature of charge. There may, however, be indications of a next level of structure around 10 TeV; there are also beautiful ideas (like, for example, those of electric and magnetic monopole duality) which may manifest at energies of the order of $\alpha^{-1} m_W (= 10$ TeV). Whether even this level of structure will give us the final clues to the nature of charge, one cannot predict. All I can say is that I am forever and continually being amazed at the depth revealed at each successive level we explore. I would like to conclude, as I did at the 1978 Stockholm Conference, with a prediction which J. R. Oppenheimer made more than twenty-five years ago and which has been fulfilled today in a manner he did not live to see. More than anything else, it expresses the faith for the future with which this greatest of decades in particle physics ends: "Physics will change even more.... If it is radical and unfamiliar... we think that the future will be only more radical and not less, only more strange and not more familiar, and that it will have its own new insights for the inquiring human spirit" (J. R. Oppenheimer, Reith Lectures, BBC, 1953).

APPENDIX A: EXAMPLES OF GRAND UNIFYING GROUPS

Appendix A is contained in Table A.1: Examples of grand unifying groups.

APPENDIX B: DOES THE GRAND PLATEAU REALLY EXIST

The following assumptions went into the derivation of the formula (*I*) in the text.

(a) $SU_L(2) \times U_{L,R}(1)$ survives intact as the electroweak symmetry group from energies $\approx \mu$ right up to M. This intact survival implies that one eschews, for example, all suggestions that (i) low-energy $SU_L(2)$ may be the diagonal sum of $SU_L^{I}(2), SU_L^{II}(2), SU_L^{III}(2)$, where I, II, III, refer to the (three?) known families; (ii) or that the $U_{L,R}(1)$ is a sum of pieces, where $U_R(1)$ may have differentially descended from a $(V+A)$-symmetric $SU_R(2)$ contained in G, or (iii) that U(1) contains a piece from a four-color symmetry $SU_c(4)$ (with lepton number as the fourth color) and with $SU_c(4)$ breaking at an intermediate mass scale to $SU_c(3) \times U_c(1)$.

(b) The second assumption which goes into the derivation of the formula above is that there are no unexpected heavy fundamental fermions, which might make $\sin^2\theta(M)$ differ from $\frac{3}{8}$—its value for the low mass fermions presently known to exist.[28]

(c) If these assumptions are relaxed, for example, for the three family group $G = [SU_F(6) \times SU_c(6)]_{L \to R}$, where $\sin^2\theta(M) = \frac{9}{28}$, we find the grand unifying mass M tumbles down to 10^6 GeV.

(d) The introduction of intermediate mass scales [for example, those connoting the breakdown of family universality, or of left-right symmetry, or of a breakdown of 4-color $SU_c(4)$ down to $SU_c(3) \times U_c(1)$] will as a rule push the magnitude of the grand unifying mass M upwards [see Salam (1979) and Shafi and Wetterich (1979)]. In order to secure a proton decay life, consonant with present empirical lower limits ($\sim 10^{30}$ years) (Learned *et al.*, 1979) this is desirable anyway. (τ_{proton} for $M \sim 10^{13}$ GeV is unacceptably low $\sim 6 \times 10^{23}$

[27]Grouping quarks (q) and leptons (l) together implies treating lepton number as the fourth color, i.e., $SU_c(3)$ extends to $SU_c(4)$ (Pati and Salam, 1974). A tribal group, by definition, contains all known families in its basic representation. Favored representations of tribal SU(11) (Georgi, 1979) and tribal SO(22) [Gell-Mann (1979) *et al.*] contain 561 and 2048 fermions!

[28]If one does not know G, one way to infer the parameter $\sin^2\theta(M)$ is from the formula:

$$\sin^2\theta(M) = \frac{\Sigma T_{3L}^2}{\Sigma Q^2} = \left[\frac{9N_q + 3N_l}{20N_q + 12N_l}\right].$$

Here N_q and N_l are the numbers of the fundamental quark and lepton SU(2) doublets (assuming these are the only multiplets that exist). If we make the further *assumption* that $N_q = N_l$ (from the requirement of anomaly cancellation between quarks and leptons) we obtain $\sin^2\theta(M) = \frac{3}{8}$. This assumption, however, is not compulsive; for example, anomalies cancel also if (heavy) mirror fermions exist (Pati *et al.*, 1975b; Pati and Salam, 1975b,c; Pati, 1975). This is the case for $[SU(6)]^4$ for which $\sin^2\theta(M) = 9/28$.

years unless there are 15 Higgs.) There is, from this point of view, an indication of there being in particle physics one or several intermediate mass scales which possibly start from around 10^4 GeV upwards. *This is the end result which I wished this Appendix to lead up to.*

REFERENCES

Akulov, V. P., D. V. Volkov, and V. A. Soroka, 1975, JETP Lett. **22**, 187.

Anderson, P. W., 1963, Phys. Rev. **130**, 439.

Arnowitt, R., P. Nath, and B. Zumino, 1975, Phys. Lett. B **56**, 81.

Ashmore, J., 1972, Nuovo Cimento Lett. **4**, 289.

Atiyah, M. F., and I. M. Singer, 1963, Bull. Am. Math. Soc. **69**, 422.

Barber, D. P., *et al.*, 1979, Phys. Rev. Lett. **43**, 830.

Barkov, L. M., 1979, in *Proceedings of the 19th International Conference on High Energy Physics* (Physical Society of Japan, Tokyo), p. 425, edited by S. Homma, M. Kawaguchi, and H. Miyazawa.

Bjorken, J. D., 1972, in the *Proceedings of the XVI International Conference on High Energy Physics* at Chicago-Batavia, edited by J. D. Jackson and A. Roberts (NAL, Batavia, Illinois), Vol. 2, p. 304.

Bludman, S., 1958, Nuovo Cimento **9**, 433.

Bollini, C., and J. Giambiagi, 1972, Nuovo Cimento B **12**, 20.

Bouchiat, C., J. Iliopoulos, and P. Meyer, 1972, Phys. Lett. B **38**, 519.

Boulware, D. G., 1970, Ann. Phys. (NY) **56**, 140.

Brandelik, R., *et al.*, 1979, Phys. Lett. B **86**, 243.

Brans, C. H., and R. H. Dicke, 1961, Phys. Rev. **124**, 925.

Brink, L., M. Gell-Mann, P. Ramond, and J. H. Schwarz, 1978, Phys. Lett. B **74**, 336.

Buras, A., J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, 1978, Nucl. Phys. B **135**, 66.

Chamseddine, A. H., and P. C. West, 1977, Nucl. Phys. B **129**, 39.

Cremmer, E., and B. Julia, 1978, Phys. Lett. B **80**, 48.

Cremmer, E., and B. Julia, 1979, Ecole Normale Supérieure preprint, LPTENS 79/6, March.

Cremmer, E., B. Julia, and J. Scherk, 1978, Phys. Lett. B **76**, 409.

Cremmer, E., and J. Scherk, 1976a, Nucl. Phys. B **103**, 399.

Cremmer, E., and J. Scherk, 1976b, Nucl. Phys. B **108**, 409.

Cremmer, E., and J. Scherk, 1976c, Nucl. Phys. B **118**, 61.

Cremmer, E., *et al.*, 1979, Nucl. Phys. B **147**, 105.

Curtwright, T. L., and P. G. O. Freund, 1979, Enrico Fermi Institute preprint EFI 79/25, April.

D'Adda, A., M. Lüscher, and P. Di Vecchia, 1978, Nucl. Phys. B **146**, 63.

Deser, S., and B. Zumino, 1976, Phys. Lett. B **62**, 335.

DeWitt, B. S., 1967a, Phys. Rev. **162**, 1195.

DeWitt, B. S., 1967b, Phys. Rev. **162**, 1239.

Dimopoulos, S., and L. Susskind, 1978, Phys. Rev. D **18**, 4500.

Dimopoulos, S., and L. Susskind, 1979, Nucl. Phys. B **155**, 237.

Dirac, P. A. M., 1931, Proc. R. Soc. (London) A **133**, 60.

Einstein, A., 1912a, Ann. Phys. (Leipzig) **38**, 355.

Einstein, A., 1912b, Ann. Phys. (Leipzig) **38**, 433.

Einstein, A., 1916, Ann. Phys. (Leipzig) **49**, 769 [English translation in *The Principle of Relativity* (Methuen, 1923), reprinted by Dover (New York), p. 35].

Elias, V., J. C. Pati, and A. Salam, 1978a, Phys. Lett. B **73**, 451.

Elias, V., J. C. Pati, and A. Salam, 1978b, Phys. Rev. Lett. **40**, 920.

Ellis, J., 1979, in *Neutrino-79*, Proceedings of the International Conference on Neutrinos, Weak Interactions, and Cosmology, June 1979, Bergen, edited by A. Haatuft and C. Jarlskog, p. 451.

Ellis, J., M. K. Gaillard, and D. V. Nanopoulos, 1979, Phys. Lett. B **80**, 360 [**82**, 464(E) (1979)].

Englert, F., and R. Brout, 1964, Phys. Rev. Lett. **13**, 321.

Englert, F., R. Brout, and M. F. Thiry, 1966, Nuovo Cimento **43**, 244.

Faddeev, L. D., and V. N. Popov, 1967, Phys. Lett. B **25**, 29.

Fayet, P., 1977, Phys. Lett. B **70**, 461.

Fayet, P., 1979, Phys. Lett. B **84**, 421.

Ferrara, S., 1979, in Proceedings of the Second Marcel Grossmann Meeting, Trieste (in preparation), and references therein.

Feynman, R. P., 1963, Acta Phys. Pol. **24**, 297.

Feynman, R. P., and M. Gell-Mann, 1958, Phys. Rev. **109**, 193.

Fradkin, E. S., and I. V. Tyutin, 1970, Phys. Rev. D **2**, 2841.

Freedman, D. Z., 1979, in *Proceedings of the 19th International Conference on High Energy Physics* (Physical Society of Japan, Tokyo), edited by S. Homma, M. Kawaguchi, and H. Miyazawa.

Freedman, D. Z., P. van Nieuwenhuizen, and S. Ferrara, 1976, Phys. Rev. D **13**, 3214.

Friedman, J. I., and V. L. Telegdi, 1957, Phys. Rev. **105**, 1681.

Fritzsch, H., and M. Gell-Mann, 1972, in *Proceedings of the XVI International Conference on High Energy Physics*, Chicago-Batavia, edited by J. D. Jackson and A. Roberts (NAL, Batavia, Illinois), Vol. 2, p. 135.

Fritzsch, H., M. Gell-Mann, and H. Leutwyler, 1973, Phys. Lett. B **47**, 365.

Fritzsch, H., and P. Minkowski, 1975, Ann. Phys. (NY) **93**, 193.

Fritzsch, H., and P. Minkowski, 1976, Nucl. Phys. B **103**, 61.

Fuller, R. W., and J. A. Wheeler, 1962, Phys. Rev. **128**, 919.

Garwin, R., L. Lederman, and M. Weinrich, 1957, Phys. Rev. **105**, 1415.

Gell-Mann, M., 1979 (unpublished).

Georgi, H., 1975, in *Particles and Fields-1974* (APS/DPF, Williamsburg), edited by C. E. Carlson (AIP, New York), p. 575.

Georgi, H., 1979, Harvard University Report No. HUTP-29/A013.

Georgi, H., and S. L. Glashow, 1974, Phys. Rev. Lett. **32**, 438.

Georgi, H., and D. V. Nanopoulos, 1979, Phys. Lett. B **82**, 392.

Georgi, H., H. R. Quinn, and S. Weinberg, 1974, Phys. Rev. Lett. **33**, 451.

Gibbons, G. W., S. W. Hawking, and M. J. Perry, 1978, Nucl. Phys. B **138**, 141.

Glashow, S. L., 1959, Nucl. Phys. **10**, 107.

Glashow, S. L., 1961, Nucl. Phys. **22**, 579.

Glashow, S., J. Iliopoulos, and L. Maiani, 1970, Phys. Rev. D **2**, 1285.

Goldstone, J., 1961, Nuovo Cimento **19**, 154.

Goldstone, J., A. Salam, and S. Weinberg, 1962, Phys. Rev. **127**, 965.

Gol'fand, Yu. A., and E. P. Likhtman, 1971, JETP Lett. **13**, 323.

Gross, D. J., and R. Jackiw, 1972, Phys. Rev. D **6**, 477.

Gross, D. J., and J. Wilczek, 1973, Phys. Rev. D **8**, 3633.

Guralnik, G. S., C. R. Hagen, and T. W. B. Kibble, 1964, Phys. Rev. Lett. **13**, 585.

Harari, H., 1979, Phys. Lett. B **86**, 83.

Hasert, F. J., *et al.*, 1973, Phys. Lett. B **46**, 138.

Hawking, S. W., 1978, Phys. Rev. D **18**, 1747.

Hawking, S. W., 1979a, in *General Relativity: An Einstein Centenary Survey* (Cambridge University, Cambridge, England).

Hawking, S. W., 1979b, "Euclidean quantum gravity," DAMTP, Univ. of Cambridge preprint.

Higgs, P. W., 1964a, Phys. Lett. **12**, 132.

Higgs, P. W., 1964b, Phys. Rev. Lett. **13**, 508.

Higgs, P. W., 1966, Phys. Rev. **145**, 1156.
Jackiw, R., 1972, in *Lectures on Current Algebra and its Applications*, by S. B. Treiman, R. Jackiw, and D. J. Gross (Princeton University, Princeton, N. J.).
Julia, B., 1979, in Proceedings of the Second Marcel Grossmann Meeting, Trieste (in preparation).
Kaluza, T., 1921, Sitzungsber. Preuss. Akad. Wiss. p. 966.
Kemmer, N., 1937, Phys. Rev. **52**, 906.
Kibble, T. W. K., 1967, Phys. Rev. **155**, 1554.
Klein, O., 1926, Z. Phys. **37**, 895.
Klein, O., 1939, "On the theory of charged fields," in *Le Magnétisme*, Proceedings of the Conference organized at the University of Strasbourg by the International Institute of Intellectual Cooperation, Paris.
Landau, L., 1957, Nucl. Phys. **3**, 127.
Learned, J., F. Reines, and A. Soni, 1979, Phys. Lett. **43**, 907.
Lee, B. W., 1972, Phys. Rev. D **5**, 823.
Lee, T. D., and C. N. Yang, 1956, Phys. Rev. **104**, 254.
Lee, T. D., and C. N. Yang, 1957, Phys. Rev. **105**, 1671.
Lee, B. W., and J. Zinn-Justin, 1972, Phys. Rev. D **5**, 3137.
Lee, B. W., and J. Zinn-Justin, 1973, Phys. Rev. D **7**, 1049.
MacDowell, S. W., and F. Mansouri, 1977, Phys. Rev. Lett. **38**, 739.
Mandelstam, S., 1968a, Phys. Rev. **175**, 1588.
Mandelstam, S., 1968b, Phys. Rev. **175**, 1604.
Marciano, W. J., 1979, Phys. Rev. D **20**, 274.
Marciano, W., and H. Pagels, 1978, Phys. Rep. C **36**, 137.
Marshak, R. E., Riazuddin, and C. P. Ryan, 1969, *Theory of Weak Interactions in Particle Physics* (Wiley-Interscience, New York).
Marshak, R. E., and E. C. G. Sudarshan, 1957, in *Proceedings of the Padua-Venice Conference on Mesons and Recently Discovered Particles* (Societa Italiana di Fisica).
Marshak, R. E., and E. C. G. Sudarshan, 1958, Phys. Rev. **109**, 1860.
Minkowski, P., 1977, University of Berne preprint, October.
Mohapatra, R. N., and J. C. Pati, 1975a, Phys. Rev. D **11**, 566.
Mohapatra, R. N., and J. C. Pati, 1975b, Phys. Rev. D **11**, 2558.
Nambu, Y., 1960, Phys. Rev. Lett. **4**, 380.
Nambu, Y., and G. Jona-Lasinio, 1961, Phys. Rev. **122**, 345.
Nanopoulos, D. V., 1979, CERN preprint TH2737, September.
Nanopoulos, D. V., and S. Weinberg, 1979, Harvard University preprint HUTP-79/A023.
Nördstrom, G., 1912, Phys. Z. **13**, 1126.
Nördstrom, G., 1913a, Ann. Phys. (Leipzig) **40**, 856.
Nördstrom, G., 1913b, Ann. Phys. (Leipzig) **42**, 533.
Nördstrom, G., 1914a, Ann. Phys. (Leipzig) **43**, 1101.
Nördstrom, G., 1914b, Phys. Z. **15**, 375.
Ogievetsky, V., and E. Sokatchev, 1978, Phys. Lett. B **79**, 222.
Pati, J. C., 1975, in *Theories and Experiments in High-Energy Physics*, edited by A. Perlmutter and S. Widmayer, Orbis Scientiae II, Coral Gables, Florida (Plenum, New York), p. 253.
Pati, J. C., and S. Rajpoot, 1978, Phys. Lett. B **79**, 65.
Pati, J. C., and A. Salam, 1973a, Phys. Rev. D **8**, 1240.
Pati, J. C., and A. Salam, 1973b, Phys. Rev. Lett. **31**, 661.
Pati, J. C., and A. Salam, 1974, Phys. Rev. D **10**, 275.
Pati, J. C., and A. Salam, 1975a, ICTP, Trieste, IC/75/106, Palermo Conference, June.
Pati, J. C., and A. Salam, 1975b, Phys. Rev. D **11**, 1137.
Pati, J. C., and A. Salam, 1975c, Phys. Rev. D **11**, 1149.
Pati, J. C., and A. Salam, 1980 (in preparation).
Pati, J. C., A. Salam, and J. Strathdee, 1975a, Phys. Lett. B **59**, 265.
Pati, J. C., A. Salam, and J. Strathdee, 1975b, Nuovo Cimento A **26**, 72.
Polyakov, A. M., 1974, JETP Lett. **20**, 194.
Rajpoot, S., and V. Elias, 1978, ICTP, Trieste, preprint IC/78/159.
Sakurai, J. J., 1958, Nuovo Cimento **7**, 1306.
Salam, A., 1957a, Nuovo Cimento **5**, 299.
Salam, A., 1957b, preprint, Imperial College, London.
Salam, A., 1968, in *Elementary Particle Theory*, Proceedings of the 8th Nobel Symposium, edited by N. Svartholm (Almqvist and Wiksell, Stockholm).
Salam, A., 1977, in *Physics and Contemporary Needs*, edited by Riazuddin (Plenum, New York), Vol. 1, p. 301.
Salam, A., 1979, in the Proceedings of the International Conference on High-Energy Physics, Geneva (European Physical Society, Cern), p. 853.
Salam, A., and J. Strathdee, 1970, Phys. Rev. D **2**, 2869.
Salam, A., and J. Strathdee, 1974a, Nucl. Phys. B **79**, 477.
Salam, A., and J. Strathdee, 1974b, Nucl. Phys. B **80**, 499.
Salam, A., and J. Strathdee, 1974c, Phys. Lett. B **51**, 353.
Salam, A., and J. Strathdee, 1978, Fortschr. Phys. **26**, 57.
Salam, A., and J. C. Ward, 1959, Nuovo Cimento **11**, 568.
Salam, A., and J. C. Ward, 1961, Nuovo Cimento **19**, 165.
Salam, A., and J. C. Ward, 1964, Phys. Lett. **13**, 168.
Scherk, J., 1979, Ecole Normale Supérieure preprint, LPTENS 79/17, September.
Schupe, M., 1979, Phys. Lett. B **86**, 87.
Schwinger, J., 1957, Ann. Phys. (NY) **2**, 407.
Shafi, Q., and C. Wetterich, 1979, Phys. Lett. B **85**, 52.
Shapiro, I. I., *et al.*, 1976, Phys. Rev. Lett. **36**, 555.
Shaw, R., 1955, "The problem of particle types and other contributions to the theory of elementary particles," Ph.D. thesis, Cambridge University (unpublished).
Siegel, W., 1977, Harvard University preprint HUTP-77/A068 (unpublished).
Slavnov, A., 1972, Theor. Math. Phys. **10**, 99.
Taylor, J. C., 1971, Nucl. Phys. B **33**, 436.
Taylor, J. G., 1977, King's College, London, preprint (unpublished).
Taylor, R. E., 1979, in *Proceedings of the 19th International Conference on High Energy Physics*, edited by S. Homma, M. Kawaguchi, and H. Miyazawa (Physical Society of Japan, Tokyo), p. 422.
't Hooft, G., 1971a, Nucl. Phys. B **33**, 173.
't Hooft, G., 1971b, Nucl. Phys. B **35**, 167.
't Hooft, G., 1974, Nucl. Phys. B **79**, 276.
't Hooft, G., and M. Veltman, 1972a, Nucl. Phys. B **44**, 189.
't Hooft, G., and M. Veltman, 1972b, Nucl. Phys. B **50**, 318.
Tiomno, J., 1956, Nuovo Cimento **1**, 226.
Tiomno, J., and J. A. Wheeler, 1949a, Rev. Mod. Phys. **21**, 144.
Tiomno, J., and J. A. Wheeler, 1949b, Rev. Mod. Phys. **21**, 153.
Toussiant, B., S. B. Treiman, F. Wilczek, and A. Zee, 1979, Phys. Rev. D **19**, 1036.
Volkov, D. V., and V. P. Akulov, 1972, JETP Lett. **16**, 438.
Weinberg, S., 1967, Phys. Rev. Lett. **27**, 1264.
Weinberg, S., 1973a, Phys. Rev. Lett. **31**, 494.
Weinberg, S., 1973b, Phys. Rev. D **8**, 4482.
Weinberg, S., 1979a, Phys. Rev. D **19**, 1277.
Weinberg, S., 1979b, Phys. Rev. Lett. **42**, 850.
Wentzel, G., 1937, Helv. Phys. Acta **10**, 108.
Wess, J., and B. Zumino, 1974, Nucl. Phys. B **70**, 39.
Wess, J., and B. Zumino, 1977, Phys. Lett. B **66**, 361.
Wheeler, J. A., 1964, in *Relativity Groups and Topology*, Proceedings of the Les Houches Summer School, 1963, edited by B. S. DeWitt and C. M. DeWitt (Gordon and Breach, New York).
Williams, J. G., *et al.*, 1976, Phys. Rev. Lett. **36**, 551.
Winter, K., 1979, in Proceedings of the International Symposium on Lepton and Photon Interactions at High Energies, Fermilab, August.
Wu, C. S., *et al.*, 1957, Phys. Rev. **105**, 1413.
Yang, C. N., and R. L. Mills, 1954, Phys. Rev. **96**, 191.
Yang, C. N., and J. Tiomno, 1950, Phys. Rev. **75**, 495.
Yoshimura, M., 1978, Phys. Rev. Lett. **41**, 381.
Zumino, B., 1975, in *Gauge Theories and Modern Field Theory*, Proceedings of a Conference held at Northeastern University, Boston, edited by R. Arnowitt and P. Nath (MIT, Cambridge, Mass.).

This page is intentionally left blank

3. Lepton-Hadron Unification

The Standard Model of the electroweak and strong interactions is a product of three groups, $G_{\rm SM} \equiv {\rm SU(3)_C} \times {\rm SU(2)}_I \times {\rm U(1)}_Y$, and hence it is characterized by three independent coupling constants $g_S \equiv g_3, g_2$ and g_1. Also, since there is a U(1) group involved, there is no charge quantization in $G_{\rm SM}$. In addition, the Standard Model has a large number of free (*ad hoc*) parameters — at least seventeen — thirteen of which are related to the Yukawa sector involving fermions. It was therefore natural to ask if this situation could be improved. For example, embedding the Standard Model group $G_{\rm SM}$ in a larger underlying group G the additional symmetries may restrict some of the arbitrary features and relate the otherwise independent parameters.

Jogesh Pati and Abdus Salam were the first to notice in 1973 that since quarks and leptons have very similar SU(2) × U(1) representation content, they may after all be very similar entities. They provided the simplest realization of the idea of quark-lepton universality by postulating that 'lepton number is the fourth colour'. This meant enlarging the ${\rm SU(3)_C}$ colour group to ${\rm SU(4)_{L+R}}$. The Pati-Salam local symmetry group is:

$$G_{\rm PS} = {\rm SU(2)_L} \times {\rm SU(2)_R} \times {\rm SU(4)_{L+R}}\,, \tag{30}$$

where the ${\rm SU(2)_{L,R}}$ group transformations act on the flavour indices. For a single family of (u, d) flavours, the corresponding multiplet is:

$$(F_e)_{\rm L,R} = \begin{pmatrix} u_r & u_y & u_b & \nu_e \\ d_r & d_y & d_b & e^- \end{pmatrix}_{\rm L,R}\,, \tag{31}$$

where the subscripts denote the Pati-Salam SU(4) quantum numbers. The multiplets $F_{\rm L,R}$ belong to $(1/2, 0, \bar{4})$ and $(0, 1/2, \bar{4})$ representations of the Pati-Salam group G_{PS}. Later, it was realized that it may be more appropriate to take the $(B - L)$ quantum number as the fourth colour. To incorporate the quark-lepton families, the group $[{\rm SU(4)}]^4$ was suggested as the simplest extension. So, apart from the known three families it stipulated a fourth one. This group is one of the simplest partial unification models containing the Standard Model symmetry group $G_{\rm SM}$ and realizing at the same time quark-lepton unification. Gauging the SU(4) symmetry, which is then broken spontaneously, involves massive gauge bosons (leptoquarks) inducing the transition $q \to l$. Hence, baryon- and lepton-number conservation cannot be exact. This line of argument had led Pati and Salam in 1973 to the proposition that the lightest of the baryons — the proton — must ultimately decay

into leptons. So, the idea of baryon-number violation in the context of gauge theories emerged. The idea of baryon instability itself predates the Pati-Salam model. In particular, Andrei Sakharov had enumerated it in his famous 1967 paper as one of the three ingredients to explain the observed baryonic asymmetry of the universe, with CP violation and statistical non-equilibrium being the other two.

It should be emphasized here that the ideas of Pati and Salam provided theoretical motivations for the experiments searching for proton decay and related phenomena such as n-$\bar{n}$ oscillation. Describing the simplest of baryon number violating interactions by the effective Lagrangian

$$\mathcal{L}_{\text{eff}} = \frac{G_B}{\sqrt{2}}(\bar{q}l)(\bar{l}l) + \text{h.c.}, \tag{32}$$

where G_B is an effective coupling constant, they predicted proton decays in the mode $p \to 3l + \pi$. This effective Lagrangian was then derived in their quark-lepton unified model discussed above. It became clear, however, that the details of the proton decay modes depend on the quark-lepton unification group and the choice of the Higgs representation needed to break the gauge symmetries. Assuming definite representations for this purpose, they emphasized a number of proton decay modes and estimated the ensuing lifetimes. Their paper and subsequent work led to a well defined experimental search programme in this field.

The minimal Pati-Salam model [Eq. (30)] is regarded as a partial unification scheme since it still has two coupling constants. This line of research subsequently led to more realistic and truly unified models for the electroweak and strong interactions — the so-called Grand Unified (GUT) Models — involving only one gauge coupling constant. Likewise, the concept of quark-lepton universality has found a more robust — perhaps also more attractive — framework provided by the GUT models, SU(5), SO(10), E_6, and others involving extensions of them to include supersymmetry. In the most popular of GUT models, namely the Georgi-Glashow SU(5), the symmetry breaking $G \to G_{\text{SM}}$ takes place directly and the Pati-Salam SU(4) representation does not emerge. However, in some others, for example the SO(10) model, the Pati-Salam $\text{SU(4)}_{\text{L+R}}$ group emerges at an intermediate step of symmetry breaking. While the Pati-Salam models are not candidates for a truly unified theory, it is entirely possible that their SU(4) group may emerge in the breaking of a larger symmetry. As discussed below, the jury is still out on the final verdict about the most viable of the GUT models.

A brief discussion of the GUT theories and their experimental implications is probably in place here. If the group G having the standard model symmetry group G_{SM} as a subgroup is simple, or if it is a direct product of identical simple groups related by discrete symmetries, then G has only one gauge coupling constant. The

first example of the embedding of the group G_{SM} in an underlying simple group — hence a truly unified model — was provided by Georgi and Glashow in 1974. The SU(5) model advocated by them puts each family of fermion fields in a reducible (5^*+10)-dimensional representation. Thus, quarks, leptons, antiquarks and antileptons appear together in these representations. There were no right-handed neutrinos in the simplest forms of the SU(5) model, and quark-lepton family replication was implemented by simply repeating the SU(5)-fermion structure for the three families. The fermion representation content of the Georgi-Glashow model is such that the gauge bosons, after spontaneous symmetry breaking, induce antiquark $\rightarrow$ lepton transitions, as opposed to the Pati-Salam gauge bosons inducing quark $\rightarrow$ lepton transitions. So, in SU(5) one expects decays of the type $p \rightarrow e^+\bar{q}q$ and $n \rightarrow e^+\bar{u}d$, leading to the decays $p \rightarrow e^+\pi^0$ and $n \rightarrow e^+\pi^-$ etc. The proton decay could then be expressed (in close analogy with the classical neutron β-decay $n \rightarrow pe^-\bar{\nu}_e$) by an effective four-Fermi interaction of strength α_5/m_X^2, where $\alpha_5 = g_5^2/4\pi$, involving the SU(5) coupling constant and the generic heavy boson mass, m_X, which is $O(m_{GUT})$. This would then give the following proton lifetime:

$$\tau_p \simeq \frac{1}{\alpha_5^2}\frac{m_X^4}{m_p^5}. \tag{33}$$

Taking $\alpha_5 \simeq \alpha$ and requiring $\tau_p > 10^{30}$ years (an experimental limit in vogue then) implies $m_X > 10^{14}$ GeV — some twelve orders of magnitude larger than m_W, m_Z. This mass scale could also be predicted by the measured values of the three gauge coupling constants g_3, g_2 and g_1. The point is that in a group such as SU(5) the generators are all part of the simple group and hence these coupling constants are all related to each other, unifying at the scale $Q^2 \geq m_X^2$

The unification of the coupling constants and the prediction of the weak angle $\sin^2\theta_W$ at low energies are some of the testable features of the Georgi-Glashow model (and variations on this theme) from low-energy precision experiments. These features deserve a brief discussion. At asymptotic energies $Q^2 \geq m_X^2$, one has the following relation for the weak angle in SU(5):

$$\sin^2\theta_W = \frac{e^2}{g_S^2} = \frac{\sum_a (t_a^3)^2}{\sum_a (q_a)^2} = \frac{\mathrm{Tr}(T^3)^2}{\mathrm{Tr}(Q^2)} = \frac{3}{8}, \tag{34}$$

where T^3 and Q are, respectively, the generators corresponding to the third component of the weak isospin and electric charge. However, because of the renormalization group (RG) equations, the corresponding angle measured at low energies is smaller due to the different Q^2-dependence of the coupling constants appearing in the numerator and denominator in the above expression. Using $\alpha_3(m_Z)_{\overline{MS}}$ and

$\alpha(m_Z)_{\overline{\rm MS}}$ (where the subscript $\overline{\rm MS}$ denotes the renormalization scheme adopted for the definition of the coupling constants) as inputs, one can predict both m_X and $\sin^2\theta_W(m_Z)_{\overline{\rm MS}}$ in GUT models such as SU(5) with no extra scales between the electroweak scale (for example, m_t) and the GUT scale m_X. Typical predictions are:

$$\begin{aligned} m_X &\simeq 2^{+2}_{-1} \times 10^{14}\ \text{GeV} \\ \sin^2\theta_W(m_Z)_{\overline{\rm MS}} &= 0.215 \pm 0.003\,. \end{aligned} \tag{35}$$

This value for $\sin^2\theta_W(m_Z)_{\overline{\rm MS}}$ differs from the presently determined value of the same quantity,

$$\sin^2\theta_W(m_Z)_{\overline{\rm MS}} = 0.2328 \pm 0.0007\,, \tag{36}$$

discussed in the section on electroweak unification, by six standard deviations. Likewise, the partial lifetime for the proton decay mode in SU(5) is predicted to be:

$$\tau(p \to e^+\pi^0) \simeq 4 \times 10^{29\pm0.7\pm1.2}\,\text{yr}\,, \tag{37}$$

which is well below the experimental bound from the IMB collaboration, $\tau(p \to e^+\pi^0) \geq 5.5 \times 10^{32}$ yr. Though these analyses are done for the SU(5) case, they are fairly representative of the general GUT models, involving a single breaking to the Standard Model.

The same mismatch between theory (single-breaking GUT models) and low-energy measurements is observed if one uses the renormalization group (RG) equations and evolves the three effective coupling constants from their measured values to the GUT scale and above. This is shown in Fig. 1. The figure labelled as the Standard Model shows that the three couplings evolved to the GUT scale do not meet at a point. This observation and the others discussed earlier very probably point to the presence of new physics between the electroweak scale $O(m_W)$ and the GUT scale $O(m_X)$. One such candidate is supersymmetry (SUSY), with a SUSY breaking scale somewhere between the electroweak scale and $O(1)$ TeV. The RG running of the coupling constants in the minimal supersymmetric extension of the Standard Model leads to solutions such that the couplings do meet at a point, with the generic superpartner mass in the range m_Z to 1 TeV, or somewhat higher. The unification scale in such extensions is higher than the value of m_X discussed above in the context of the SU(5) model. This would imply a longer lifetime for the proton, evading the present experimental bound. The details of the decay modes in such extensions are also different.

These and related issues will come under more precise experimental and theoretical scrutiny in not too distant a future. Improved experiments are on their way which would pin down the mechanism responsible for the apparent solar neutrino

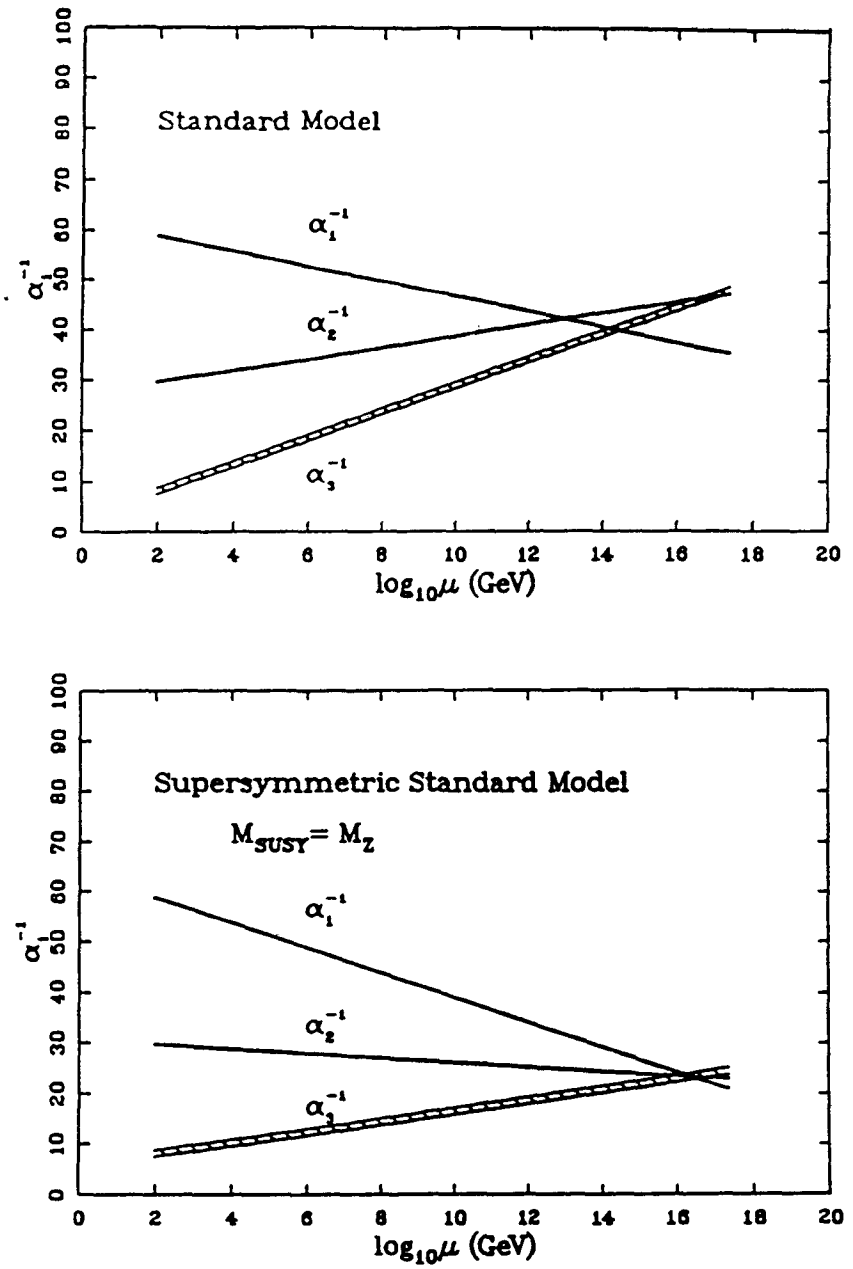


Figure 1: Running of the three gauge couplings, assuming the Standard Model and its minimal supersymmetric extension. (Source: Paul Langacker, in *Proceedings of the Neutrino Telescope*, Venice, 1992.)

deficit. Along the same lines, laboratory oscillation experiments to measure neutrino masses and mixings that are of interest for the solar neutrino problem are being set up. These experiments will establish the existence of, or put a firm limit on, lepton number violation. Likewise, improved sensitivity to the proton lifetime is anticipated in future experiments, from which one expects to learn about the scale and texture of baryon number violating processes. These experimental devel-

opments are expected to disentangle the physics between the electroweak and GUT scales. It looks almost certain that the Standard Model will have to be extended. A good candidate for the extension is supersymmetry, and the search for supersymmetric particles at present and planned accelerators is one of the foremost research goals. At present, both the SUSY extension of the Standard Model and its eventual embeddings in GUT scenarios are far from being obvious. While the quest for the final theory of elementary particles continues, the idea that quarks and leptons can be treated on an equal footing, and that both lepton and baryon number violations indeed take place in such unified theories, which Salam and Pati proposed in 1973, is now an integral part of GUT models.

Reprinted from:
PHYSICAL REVIEW D VOLUME 8, NUMBER 4 15 AUGUST 1973

Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions*

Jogesh C. Pati†
Department of Physics and Astronomy, University of Maryland, College Park, Maryland

Abdus Salam
International Centre for Theoretical Physics, Miramare, Trieste, Italy
and Imperial College, London
(Received 5 February 1973)

An attempt is made to unify the fundamental hadrons and leptons into a common irreducible representation F of the same symmetry group G and to generate a gauge theory of strong, electromagnetic, and weak interactions. Based on certain constraints from the hadronic side, it is proposed that the group G is $SU(4')\times SU(4'')$, which contains a Han-Nambu-type $SU(3')\times SU(3'')$ group for the hadronic symmetry, and that the representation F is $(4,4^*)$. There exist four possible choices for the lepton number L and accordingly four possible assignments of the hadrons and leptons within the $(4,4^*)$. Two of these require nine Han-Nambu-type quarks, three "charmed" quarks, and the observed quartet of leptons. The other two also require the nine Han-Nambu quarks, plus heavy leptons in addition to observed leptons and only one or no "charmed" quark. One of the above four assignments is found to be suitable to generate a gauge theory of the weak, electromagnetic, and $SU(3'')$ gluonlike strong interactions from a selection of the gauges permitted by the model. The resulting gauge symmetry is $SU(2')_L\times U(1)\times SU(3'')_{L+R}$. The scheme of all three interactions is found to be free from Adler-Bell-Jackiw anomalies. The normal strong interactions arise effectively as a consequence of the strong gauges $SU(3'')_{L+R}$. Masses for the gauge bosons and fermions are generated suitably by a set of 14 complex Higgs fields. The neutral neutrino and $\Delta S=0$ hadron currents have essentially the same strength in the present model as in other $SU(2)_L\times U(1)$ theories. The mixing of strong- and weak-gauge bosons (a necessary feature of the model) leads to parity-violating nonleptonic amplitudes, which may be observable depending upon the strength of $SU(3'')$ symmetry breaking. The familiar hadron symmetries such as $SU(3')$ and chiral $SU(3')_L\times SU(3')_R$ are broken only by quark mass terms and by the electromagnetic and weak interactions, not by the strong interactions. The difficulties associated with generating gauge interactions in the remaining three assignments are discussed in Appendix A. Certain remarks are made on the question of proton and quark stability in these three schemes.

I. INTRODUCTION

Several past attempts[1] including some of the recently proposed schemes of unification[2,3] of the weak and electromagnetic interactions have emphasized the parallelism between the hadrons and leptons in the sense that both types of matter belong to similar *(but distinct)* multiplets of the same group structure[1,4] [e.g., $SU(2)\times U(1)$ or $SU(4)$, etc.]. As a further step toward a unified theory of matter, one may inquire whether the fundamental hadrons and the leptons could be considered as belonging to the *same* irreducible representation of a *common* symmetry group for all matter. It of course goes without saying that such a supersymmetry group has to be badly broken to correspond to the huge mass gap between the hadrons and the leptons and to their asymmetric response toward the strong interactions.

Nevertheless, if we postulate that such a symmetry group is the basis for the existence of hadrons *and* leptons, it may in the first place lead to a tighter classification scheme for the leptons not available at present. This could provide a rationale for the existence of known leptons and also throw some light on further leptons to be expected. The existence of the lepton number L (or baryon number B) would derive an interpretation similar to that of I_3 and Y, being part of a hierarchy of higher symmetry rather than due to an *ad hoc* $U(1)$ symmetry. Such a combined symmetry is also expected to provide a well-defined basis for hadron-lepton universality in the weak and electromagnetic currents.

In the second place, the full dynamical content of such a supersymmetry group may provide new insights. For example, a *full* gauge theory of interactions based on such a symmetry group may lead to a unified treatment of the strong as well as weak electromagnetic interactions in the sense that a single coupling constant may govern the non-Abelian part of the interactions. The empir-

ical difference between the value of the electromagnetic charge and the strong interaction coupling constant may, in such a scheme, be expected to arise from differing charge renormalizations produced by the spontaneous symmetry breaking mechanisms, which are a necessary feature of such theories. A related consequence is the possibility of a whole new class of interactions involving gauge mesons which carry baryon and lepton number. The absence of such interactions in the present energy domain may be attributable to superheavy masses of such gauge mesons, consistent with current ideas of spontaneous symmetry breaking.

With these to serve as motivation, we attempt to make a beginning toward a unified treatment of the hadrons and leptons. In Sec. II we discuss the question of the choice of the symmetry group G and the representation F for fundamental hadronic and leptonic matter. Based on certain constraints from the hadron's side, it is proposed that the relevant group for the *classification* of fundamental hadrons and leptons is $SU(4') \times SU(4'')$, which contains a Han-Nambu-type[5] $SU(3') \times SU(3'')$ group for the hadronic symmetry, and that the representation F is $(4, 4^*)$. It is then pointed out that there exist four distinct possibilities for the assignment of the leptons within the $(4, 4^*)$. Two of these require nine Han-Nambu-type "normal" quarks, three "charmed" quarks, and just the observed quartet of leptons (but no heavy leptons). The other two also require nine Han-Nambu quarks, but they require extra leptons and correspondingly one or no charmed quark.

In Sec. III, we construct a gauge theory of weak, electromagnetic, and $SU(3'')$ gluonlike strong interactions for *one* of the four assignments mentioned above. We restrict ourselves in this note, however, to only a selected set of gauges so as to reproduce primarily the conventional phenomena of interactions for this assignment. (The other three assignments seem to be unsuitable for a variety of reasons discussed in Appendix A.) The symmetry group for the gauge interactions thus obtained is $SU(2')_L \times U(1) \times SU(3'')_{L+R}$. *The scheme, including the strong interactions, is found to be free from Adler-Bell-Jackiw anomalies.* The existence of the two SU(4) degrees of freedom for the $(4, 4^*)$ is found to be essential in realizing this scheme.

The model thus generated is conservative in the sense that it does not exploit all the gauge symmetries permitted by the model. By the same token, no attempt is made to build in universality of the strong, weak, and electromagnetic interactions in the sense mentioned before. We hope to return to these considerations in subsequent work.

The question of generating masses for the gauge mesons and the $(4, 4^*)$ fermions via spontaneous symmetry breaking is briefly discussed in Sec. IV. A mass matrix with desired properties is constructed with 14 complex Higgs fields. It is pointed out that the mass matrix must necessarily mix some of the strong-gauge mesons with the weak ones in order to generate the photon. This may lead to observable effects in parity-violating nuclear transitions depending upon a few factors. The neutral neutrino and $\Delta S = 0$ hadron currents are found to have essentially the same strength in the present model as in other $SU(2)_L \times U(1)$ theories[3,4] of weak and electromagnetic interactions. Remarks on the nature of various symmetries in the model are made in Sec. IV. Section V contains a summary of a few peculiar features of the scheme. The difficulties associated with generating gauge interactions in three out of the four lepton assignments are discussed in Appendix A. The possibility of baryon- and lepton-number–violating transitions in these cases and their implications on quark and proton stability are briefly mentioned. In Appendix B we consider the consequences of assuming that the idea of Cabibbo rotation applies to the (e, μ) leptons as it does to the $(\mathfrak{N}, \lambda)$ hadrons. It is remarked that the two-neutrino experiment (rather than absence of $\mu \to e + \gamma$ decay) would provide a test of whether the universality of hadrons and leptons should be carried this far.

II. CHOICE OF THE SYMMETRY GROUP AND THE BASIC REPRESENTATION

We shall start with the basic assumption that the representation F of the fundamental entities consists of quarks and at least the known leptons. This excludes the hypothesis that the observed leptons may be composites of something even more fundamental.[6] So far as quarks are concerned, it appears that there are several considerations, which favor nine degrees of freedom for the quarks rather than three. Apart from integral charges, these are (a) the experimental value for the magnitude of the $\pi^0 \to 2\gamma$ amplitude[7]; (b) the desirability of having Fermi statistics for the 56-plet of observed baryons, with quarks in relative s states; and (c) the question of the $\Delta T = \frac{1}{2}$ rule[8] for the nonleptonic decays. To be more specific, we will assume[9] that the basic hadrons are the Han-Nambu[5] quarks transforming as $(3, 3^*)$ under the $SU(3') \times SU(3'')$ group; for our discussion the familiar SU(3) group may be identified with $SU(3')$. The quarks will carry integral charges if we assume

$$Q=(I_3'+\tfrac{1}{2}Y')+(I_3''+\tfrac{1}{2}Y''). \tag{1}$$

The low-lying baryons and mesons are assumed to be singlets under SU(3″), thus they have $I_3''=Y''=0$.

Clearly the supergroup G we are seeking must contain the hadronic SU(3′)×SU(3″) group as a subgroup while the lepton number must be a diagonal generator within G outside of SU(3′)×SU(3″). The basic representation F of the supergroup G must contain at least the Han-Nambu nonet of hadrons and the observed quartet of leptons. It is easy to convince oneself that the simplest group structure[10] meeting these constraints is provided by

$$G=\mathrm{SU}(4')\times\mathrm{SU}(4''), \tag{2}$$

with the fundamental fermion (hadron and lepton) representation being given by

$$F=(4,4^*). \tag{3}$$

Denote the components of F by $\psi_{(\alpha,i)}$; the group SU(4′) acts on the index $\alpha=\mathcal{P},\mathfrak{N},\lambda,\chi$, and the group SU(4″) acts on the index $i=a,b,c,d$. The traceless matrices I_3', Y', and C' for the diagonal generators of SU(4′) with respect to the indices α may then be defined by

$$I_3'=\tfrac{1}{2}\begin{pmatrix}\mathcal{P} & \mathfrak{N} & \lambda & \chi\\ 1 & & & \\ & -1 & & \\ & & 0 & \\ & & & 0\end{pmatrix},\qquad Y'=\tfrac{1}{3}\begin{pmatrix}\mathcal{P} & \mathfrak{N} & \lambda & \chi\\ 1 & & & \\ & 1 & & \\ & & -2 & \\ & & & 0\end{pmatrix},$$

$$C'=\tfrac{1}{4}\begin{pmatrix}\mathcal{P} & \mathfrak{N} & \lambda & \chi\\ 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -3\end{pmatrix}, \tag{4}$$

where C' is the charm quantum number of SU(4′). The matrices I_3'', Y'', and C'' for the diagonal generators of SU(4″) *(with respect to the indices i)* may be defined similarly. They are *negatives* of I_3', Y', and C', since they correspond to the 4* representation. Thus,

$$(I_3'')_i=-(I_3')_\alpha,\qquad (Y'')_i=-(Y')_\alpha$$

and

$$(C'')_i=-(C')_\alpha. \tag{5}$$

A suitable extension of the charge formula (1), which gives integer charges to all constituent particles, is given by

$$Q=(I_3'+\tfrac{1}{2}Y'-\tfrac{2}{3}C')+(I_3''+\tfrac{1}{2}Y''-\tfrac{2}{3}C''). \tag{6}$$

One may exhibit ψ by a (4×4) array, the superscript on each component designating the charge:

$$\Psi=\begin{bmatrix}\mathcal{P}_a^0 & \mathcal{P}_b^+ & \mathcal{P}_c^+ & \mathcal{P}_d^0\\ \mathfrak{N}_a^- & \mathfrak{N}_b^0 & \mathfrak{N}_c^0 & \mathfrak{N}_d^-\\ \lambda_a^- & \lambda_b^0 & \lambda_c^0 & \lambda_d^-\\ \chi_a^0 & \chi_b^+ & \chi_c^+ & \chi_d^0\end{bmatrix}. \tag{7}$$

The 3×3 top left array denotes the Han-Nambu nonet of hadronic quarks, which we denote subsequently by $\mathcal{K}$.

There appear, in general, four distinct possibilities[11] for the choice of the lepton number L:

$$\begin{aligned}L&=(C''+\tfrac{1}{4}) && \text{(A)}\\ &=(C'-\tfrac{1}{4}) && \text{(B)}\\ &=(C''+\tfrac{1}{4})+(C'-\tfrac{1}{4}) && \text{(C)}\\ &=(C''+\tfrac{1}{4})-(C'-\tfrac{1}{4}). && \text{(D)}\end{aligned} \tag{8}$$

Accordingly there are four possible assignments of the hadrons and leptons in the (4, 4*). These are

$$\Psi_A=\begin{bmatrix}\mathcal{K}\ (L=0) & \begin{matrix}\nu_e\\ e^-\\ \mu^-\end{matrix}\ \Big\}(L=1)\\ \begin{matrix}h^0\ h^+\ h'^+\\ (L=0)\end{matrix} & \nu_\mu\end{bmatrix},$$

$$\Psi_B=\begin{bmatrix}\mathcal{K}\ (L=0) & \begin{matrix}H^0\\ H^-\\ H'^-\end{matrix}\ \Big\}(L=0)\\ \begin{matrix}\bar\nu_e\ e^+\ \mu^+\\ (L=-1)\end{matrix} & \bar\nu_\mu\end{bmatrix},$$

$$\Psi_C=\begin{bmatrix}\mathcal{K}\ (L=0) & \begin{matrix}\nu_\mu\\ \mu^-\\ M^-\end{matrix}\ \Big\}(L=1)\\ \begin{matrix}\bar\nu_e\ e^+\ E^+\\ (L=-1)\end{matrix} & \begin{matrix}R^0\\ (L=0)\end{matrix}\end{bmatrix},$$

$$\Psi_D=\begin{bmatrix}\mathcal{K}\ (L=0) & \begin{matrix}\nu_\mu\\ \mu^-\\ M^-\end{matrix}\ \Big\}(L=1)\\ \begin{matrix}\bar\nu_e\ e^+\ E^+\\ (L=-1)\end{matrix} & \begin{matrix}\delta^0\\ (L=2)\end{matrix}\end{bmatrix}. \tag{9}$$

In each case $\mathcal{K}$ stands for the 3×3 array of the Han-Nambu quarks. Note that:

(1) All four schemes accommodate at least the

known quartet of leptons and allow for the distinction between ν_e and ν_μ, and e and μ via SU(4′) and/or SU(4″) quantum numbers.

(2) Schemes (A) and (B) contain *just the observed quartet of leptons* and no further leptons, while (C) and (D) contain new leptons in addition.

(3) Schemes (A) and (B) look so similar that one may not regard them as distinct. However, once we choose the convention that the low-lying hadrons are singlets with respect to SU(3″) and nonsinglet in general with respect to SU(3′), then (A) and (B) lead to physically distinct schemes of gauge interactions (compare Sec. III and Appendix A).

(4) In addition to the basic Han-Nambu nonet, schemes (A) and (B) consist of three charmed hadrons, called (h^0,h^+,h'^+) and (H^0,H^-,H'^-), respectively, while scheme (C) contains one such charmed hadron, the R^0. These objects are presumably much heavier than their uncharmed partners and do not play any role in the formation of the low-lying hadrons.

(5) In each of these four schemes, one may introduce a U(1) symmetry to extend the SU(4′) or the SU(4″) to a U(4′) or U(4″) group, respectively. The corresponding U(1) symmetry may be identified with *fermion-number conservation*. One may assign the fermion number of the basic 16-plet $\psi_{(\alpha,i)}$ to be +1 and their antiparticles to be −1. The theory permits separate conservation of the fermion number F and the lepton number L. All the traditional selection rules usually following from baryon-number and lepton-number conservations are now given by the conservations of the fermion number and the lepton number. Baryon number may, of course, be defined in these schemes in terms[12] of F and L.

In Sec. III we present a gauge theory of strong, electromagnetic, and weak interactions in scheme (A) only. The difficulties encountered in generating such interactions in the other three schemes are discussed in Appendix A.

III. GAUGE INTERACTIONS IN SCHEME (A)

A. The Weak Gauges

Since the hadrons and leptons in scheme (A) have SU(4′) quartet structures separately, we can obtain the Cabibbo suppression for the $|\Delta S|=1$ charged currents and avoid the $|\Delta S|=1$ neutral currents by following the scheme of Glashow, Iliopoulos, and Maiani.[4] Thus, to start with, we *assume* that the hadronic quarks $\mathfrak{N}_i$ and λ_i (with $i=a,b,c$), which enter directly into the currents, *are not physical*, in the sense that the mass matrix is not diagonal with respect to them; they are related to the physical objects $\tilde{\mathfrak{N}}_i$ and $\tilde{\lambda}_i$ $(i=a,b,c)$ by a Cabibbo rotation of angle θ:

$$\begin{pmatrix}\mathfrak{N}_i\\ \lambda_i\end{pmatrix}=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\tilde{\mathfrak{N}}_i\\ \tilde{\lambda}_i\end{pmatrix}\quad (i=a,b,c). \tag{10}$$

Following Ref. 4, the weak interactions may be generated by a $SU(2')_L$-gauge group with generators given by the sum of the SU(2′) generators acting on the *left-handed* (p,n) and (λ,χ) indices, respectively. These are[13]

$$\begin{aligned}K_1^L&=(F_1'+F_{13}')_L\,,\\ K_2^L&=(F_2'-F_{14}')_L\,,\\ K_3^L&=(F_3'+\tfrac{1}{2}Y'-\tfrac{2}{3}C')_L\,,\end{aligned} \tag{11}$$

where

$$Y'=(\tfrac{4}{3})^{1/2}F_8' \tag{12}$$

and

$$C'=(\tfrac{3}{2})^{1/2}F_{15}'\,.$$

Using the representation of K_j^L with respect to the left-handed basis $\alpha_L^i\equiv(\mathcal{P}_i,\mathfrak{N}_i,\lambda_i,\chi_i)_L$ as

$$[K_j^L]=\tfrac{1}{2}\begin{matrix}\\ \alpha_L^i\downarrow\end{matrix}\begin{bmatrix}\tau_j & 0\\ 0 & \bar{\tau}_j\end{bmatrix}\quad (i=a,b,c,d;\ j=1,2,3), \tag{13}$$

where τ_j are the Pauli matrices and $\bar{\tau}_1=\tau_1$, $\bar{\tau}_{2,3}=-\tau_{2,3}$, the corresponding fermion–gauge-boson interaction is given by

$$\mathcal{L}_1=g\sum_{j=1}^{3}\sum_{i=a,b,c,d}(\bar{\mathcal{P}}_i\,\bar{\mathfrak{N}}_i\,\bar{\lambda}_i\,\bar{\chi}_i)[K_j^L]\gamma_\mu[\tfrac{1}{2}(1+\gamma_5)]\times\begin{pmatrix}\mathcal{P}_i\\ \mathfrak{N}_i\\ \lambda_i\\ \chi_i\end{pmatrix}W_\mu^j\,. \tag{14}$$

Note that the right-handed fields $(\Psi_A)_R$ are being treated as singlets under the $SU(2')_L$ gauge group. As in Ref. 4, the interaction (14) suppresses $|\Delta S|=1$ neutral currents and has Cabibbo form for charged currents.

B. The Strong Interactions

The strong interactions (by definition) do not involve the leptons. We shall assume that they respect at least the SU(3′)×SU(3″) symmetry, that they are generated via pure vector and/or axial-vector gauges in order to conserve parity, and that they do not introduce anomalies by themselves or in the presence of other interactions.

All these requirements are met easily [in scheme (A)] if we use an SU(3″) octet of vector

gauges to generate the basic (gluonlike) strong interactions. Denoting the corresponding gauge mesons by V_μ^m ($m=1,\ldots,8$), their interactions with fermions are given by

$$\mathcal{L}_2 = f \sum_{m=1}^{8} \sum_{\alpha=\mathcal{P},\overline{\mathcal{R}},\lambda,\chi} \left[(\bar{\alpha}_a \bar{\alpha}_b \bar{\alpha}_c)(\tfrac{1}{2}\lambda_m)\gamma_\mu \begin{pmatrix} \alpha_a \\ \alpha_b \\ \alpha_c \end{pmatrix} \right] V_\mu^m . \tag{15}$$

The gauge mesons transform as (1, 8) under the SU(4′)×SU(3″) group. The λ_m's are the appropriate Gell-Mann matrices for the 3* representation. The interactions as introduced above are, of course, anomaly-free by themselves, since they are vectorial in nature. Remarkably enough, they are found to be anomaly-free also in the presence of the weak $SU(2')_L$ gauge interactions and the U(1) gauge interaction [see Eq. (18)] yet to be introduced to generate electromagnetism. Instead of the SU(3″) octet gauge mesons, however, if we had used a singlet gluon gauge vector meson,[14] coupled symmetrically to the twelve hadrons [in scheme (A)] to generate the strong interactions, the resulting scheme together with the U(1)-gauge interaction needed to generate electromagnetism would have anomalies. In this sense then, the existence of the two (see Ref. 14) SU(4) degrees of freedom in our model seems to play an *essential role* in generating an anomaly-free gauge theory of the strong, weak, and electromagnetic interactions.

C. Electromagnetism: A New U(1) Gauge

Note that the electric charge operator given by Eq. (6) can be written [with usual notations: $I_3' = (I_3')_L + (I_3')_R \equiv (I_3')_{L+R}$, etc.] as

$$Q = [I_3' + \tfrac{1}{2}Y' - \tfrac{2}{3}C']_L + [I_3'' + \tfrac{1}{2}Y'']_{L+R} + \{(I_3' + \tfrac{1}{2}Y' - \tfrac{2}{3}C')_R - \tfrac{2}{3}C''_{L+R}\}. \tag{16}$$

We have so far introduced $SU(2')_L$ and $SU(3'')_{L+R}$ gauges, which generate currents corresponding to the expressions in the square brackets in Eq. (16). No such gauge has been introduced to generate the current corresponding to the piece of the electric charge in the curly bracket in (16). This consists of a part $(I_3' + \frac{1}{2}Y' - \frac{2}{3}C')_R$ belonging to the $SU(2')_R$ group [compare with generators of the $SU(2')_L$ group as in Eq. (11)] and a part $(-\frac{2}{3}C''_{L+R})$ belonging to the SU(4″) group. The corresponding generators commute with the gauge generators already introduced, so that the simplest (though perhaps not the most elegant) possibility to build electromagnetism is to introduce a new U(1) gauge interaction defined by the generator

$$\tilde{Y} = (I_3' + \tfrac{1}{2}Y' - \tfrac{2}{3}C')_R - \tfrac{2}{3}C''_{L+R}. \tag{17}$$

The corresponding gauge meson U^0 interacts with the fermions in the following manner:

$$\mathcal{L}_3 = g' U_\mu^0 \left\{ \tfrac{1}{2} \sum_{i=a,b,c,d} [(\bar{\mathcal{P}}_i \mathcal{P}_i - \bar{\mathcal{R}}_i \mathcal{R}_i)_R + (\bar{\chi}_i \chi_i - \bar{\lambda}_i \lambda_i)_R] + \tfrac{1}{6} \sum_{\alpha=\mathcal{P},\overline{\mathcal{R}},\lambda,\chi} (\bar{\alpha}_a \alpha_a + \bar{\alpha}_b \alpha_b + \bar{\alpha}_c \alpha_c - 3\bar{\alpha}_d \alpha_d)_{L+R} \right\}, \tag{18}$$

where the subscripts R and $(L+R)$ on the currents are defined by the conventional notation

$$\bar{\psi}_1 \gamma_\mu [\tfrac{1}{2}(1-\gamma_5)] \psi_2 \equiv (\bar{\psi}_1 \psi_2)_R$$

and (19)

$$\bar{\psi}_1 \gamma_\mu \psi_2 \equiv (\bar{\psi}_1 \psi_2)_{L+R}.$$

The total weak, electromagnetic, and strong interaction of the theory (not including the Higgs bosons to be introduced) is

$$\mathcal{L}' = \mathcal{L}_1 + \mathcal{L}_3 + \mathcal{L}_2 \tag{20}$$

and the corresponding gauge group is given by

$$G' = SU(2')_L \times U(1) \times SU(3'')_{L+R}. \tag{21}$$

The total number of neutral gauge fields in the theory is four: W^3, V^3, V^8, and U^0. The mass matrix arising through spontaneous breaking will in general induce mixing between these fields, which in turn will lead to certain orthogonal combinations of these fields to correspond to the physical particles. One such combination is the massless photon (coupled to the conserved electric-charge gauge). The complexion of the others will depend upon the nature of the mass matrix, which is discussed briefly in Sec. IV.

IV. PARTICLES AND MASSES

Below we demonstrate a possible scheme[15] of spontaneous symmetry breaking for generating masses of the particles. We sketch the main ideas only briefly, since these are simple extensions of standard ideas in the subject.

All the gauge mesons and the fermions can be given masses appropriately by introducing two sets of scalar fields, i.e.,

(a) a set of 12 complex fields σ_α^i transforming as $(2+2, 1, 3^*)$ under the gauge group G', where $i = a, b, c$ corresponds to SU(3″) index and $\alpha = 1, 2, 3, 4$

corresponds to $SU(2')_L$ indices, plus

(b) a second set of two complex fields $\phi = \binom{\phi^+}{\phi^0}$ transforming as (2, 1, 1) under G'.

We assume that the self-interactions among the scalar fields may be arranged[16] such that their vacuum expectation values are given by

$$\langle \sigma_\alpha^i \rangle_0 = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \\ 0 & 0 & 0 \end{pmatrix} \tag{22}$$

and

$$\langle \phi \rangle_0 = \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{23}$$

where σ and λ are real. The three nonzero elements in $\langle \sigma_\alpha^i \rangle$ are taken to be the same in order to obtain (approximately) equal masses for the $SU(3'')$ octet of gauge mesons.

The masses of the gauge mesons are found by substituting the vacuum expectation values for the corresponding fields in the expressions

$$\sum_{(i,\alpha)} |\partial_\mu \sigma_\alpha^i - if(\tfrac{1}{2}\vec{\lambda}(\sigma)\cdot\vec{V}_\mu)_{ij}\sigma_\alpha^j - ig(\vec{t}(\sigma)\cdot\vec{W}_\mu)_{\alpha\beta}\sigma_\beta^i - ig'(t^0(\sigma)U_\mu^0)_{(i,\alpha)(j,\beta)}\sigma_\beta^j|^2$$
$$+ \sum_{\xi=(+,0)} |\partial_\mu \phi^\xi - ig(\vec{t}(\phi)\cdot\vec{W}_\mu)_{\xi\eta}\phi^\eta - ig'(t^0(\phi)U_\mu)_{\xi\eta}\phi^\eta|^2, \tag{24}$$

where $\vec{\lambda}(\sigma)$ ar the Gell-Mann matrices for the 3* representation and

$$(t_K(\sigma))_{\alpha\beta} = \tfrac{1}{2}\begin{pmatrix} \tau_K & 0 \\ 0 & \bar{\tau}_K \end{pmatrix}_{\alpha\beta} (\bar{\tau}_1 = \tau_1,\ \bar{\tau}_{2,3} = -\tau_{2,3}),$$
$$(t^0(\sigma))_{(i,\alpha),(j,\beta)} = \tfrac{1}{6}\delta_{ij}\delta_{\alpha\beta},$$
$$(\vec{t}(\phi))_{\xi\eta} = (\tfrac{1}{2}\vec{\tau})_{\xi\eta}, \tag{25}$$

and

$$(t^0(\phi))_{\xi\eta} = \tfrac{1}{2}\delta_{\xi\eta}.$$

This scheme leads to the following eigenstates and masses after necessary diagonalization of the mixed fields:

$$V_\rho^\pm = V_\rho^\pm \cos\delta + W^\pm \sin\delta, \quad m(V_\rho^\pm) = (\tfrac{1}{2})^{1/2} f\sigma + O(\sigma/\lambda);$$
$$V_{K^*}^\pm = (\tfrac{1}{2})^{1/2}(V^4 \mp iV^5), \quad m(V_{K^*}^\pm) = (\tfrac{1}{2})^{1/2} f\sigma;$$
$$(V_{K^*}^0, \bar{V}_{K^*}^0) = (\tfrac{1}{2})^{1/2}(V^6 \mp iV^7), \quad m(V_{K^*}^0) = (\tfrac{1}{2})^{1/2} f\sigma;$$
$$\bar{W}^\pm = W^\pm \cos\delta - V_\rho^\pm \sin\delta, \quad m(\bar{W}) = (\tfrac{1}{2})^{1/2} g\lambda + O(\sigma/\lambda);$$
$$A = \frac{f(g'W_3 + gU^0) + gg'[V_3 + (\frac{1}{3})^{1/2} V_8]}{[f^2(g^2+g'^2) + g^2 g'^2(\frac{4}{3})]^{1/2}}, \quad m_A = 0; \tag{26}$$
$$V_{Y^0} = \tfrac{1}{2}(V_3 - \sqrt{3}\, V_8), \quad m(V_{Y^0}) = (\tfrac{1}{2})^{1/2} f\sigma;$$
$$V_{Y'^0} = \frac{gg'(g'W_3 + gU^0) - (\frac{3}{4})f(g^2+g'^2)[V_3 + (\frac{1}{3})^{1/2} V_8] + O(\sigma/\lambda)}{\{g^2 g'^2 (g^2+g'^2)^2 + \frac{3}{4} f^2 (g^2+g'^2)\}^{1/2}}, \quad m(V_{Y'^0}) = (\tfrac{1}{2})^{1/2} f\sigma + O(\sigma/\lambda);$$
$$Z^0 = \frac{gW_3 - g'U^0}{(g^2+g'^2)^{1/2}} + O(\sigma/\lambda), \quad m(Z^0) = (g^2+g'^2)^{1/2}(\tfrac{1}{2}\lambda) + O(\sigma/\lambda);$$

where

$$\tan 2\delta = -2\sqrt{2}\left(\frac{\sigma^2}{\lambda^2}\right)\left(\frac{f}{g}\right)\left[1 - \left(\frac{f^2}{g^2}\right)\left(\frac{\sigma^2}{\lambda^2}\right)\right]^{-1}. \tag{27}$$

We could have exhibited the exact expressions for the diagonal fields. These, however, are complicated and add nothing to the understanding. Thus, in the above equations we have consistently assumed σ/λ to be a small number of order 10^{-2} to 10^{-3} and neglected corrections to this order. This is consistent with the assumption of a large W mass $m_W \sim (\frac{1}{2})^{1/2} g\lambda \gtrsim 50$ BeV together with the strong-gauge-meson masses $m(V) \sim (\frac{1}{2})^{1/2} f\sigma$ of around 3 to 5 BeV.[17] Taking g/f to be nearly equal to the ratio of the electromagnetic charge to typical strong interaction coupling constants (i.e., g/f lying between $\frac{1}{10}$ and $\frac{1}{30}$), we obtain[18]

$$\left(\frac{\sigma}{\lambda}\right) \simeq \left[\frac{m(V)}{m(W)}\right]\left(\frac{g}{f}\right)$$
$$\simeq (10^{-2} \text{ to } 10^{-3}). \tag{28}$$

The following points may be noted:

(1) The exchange of $V_\rho^\pm$ leads to a $|\Delta S| = 0, 1$ parity-violating amplitude for processes involving

hadrons, which is of order $(\sin\delta)(g/f)\sim(\sigma^2/\lambda^2)$ relative to the parity-conserving amplitude. Such an amplitude however involves a pure SU(3″) non-singlet transition operator, and therefore vanishes by SU(3″) symmetry for processes involving low-lying hadrons, which are supposed to be SU(3″) singlets. If SU(3″) is not a very strict symmetry (see remark at the end of this section), then for $\sigma^2/\lambda^2\sim10^{-5}$, the above contribution may still be comparable to the normal weak-interaction contribution and could serve as a distinguishing feature[19] of the model. It should be noted that the additional parity-violating amplitude for hadrons (mentioned above) is a *necessary* feature of our model, since the strong- and weak-gauge mesons must be mixed through the mass matrix in order to generate the photon.

(2) The strong interactions are mediated by the following octet of gauge mesons: $V_\rho^\pm$, $V_{K^*}^\pm$, $\overline{V}_{K^*}^0$, $V_{K^*}^0$, V_{χ^0}, and V_{γ^0}, all with nearly equal mass $(\frac{1}{2})^{1/2}f\sigma$, which is necessary to preserve SU(3″) symmetry. It should be stressed that these mesons are primarily singlets in the conventional SU(3′), but octet under SU(3″); thus they *cannot be identified* with any of the known vector mesons. We have chosen the subscripts $\rho, K^*, \ldots$, etc., for notational simplicity.

(3) The V_{γ^0} interaction, despite the relatively large component of W_3 and U^0, is parity conserving at least to order σ^2/λ^2 since the combination $(g'W_3+gU^0)$ is coupled to pure vector current. Z^0 interaction is, of course, parity violating.

(4) The strengths of neutral current processes[20] such as $\nu_e+\mu\to\nu_e+\mu$, $\nu_\mu+e\to\nu_\mu+e$, and $\nu+p\to\nu+p$, etc., are the same (to order σ/λ) in the present model as in other $SU(2)_L\times U(1)$ theories[3,4] without the strong gauge mesons. The presence of strong interactions in our model and the mass mixing does not alter the situation.

The Masses of the Fermions. The fermion masses can be generated appropriately by introducing the following gauge-invariant interaction of the ϕ field[21] with the fermions:

$$\mathcal{L}(\psi,\phi)=\sum_{i=a,b,c,d}\left[(G_{\mathcal{P}_i}\overline{\mathcal{P}}_{iR}\xi+G_{\mathfrak{N}_i}\overline{\mathfrak{N}}_{iR}\overline{\phi}+G_{\mathfrak{N}_i\lambda_i}\overline{\lambda}_{iR}\overline{\phi})\begin{pmatrix}\mathcal{P}_i\\ \mathfrak{N}_i\end{pmatrix}_L+(G'_{\mathfrak{N}_i\lambda_i}\overline{\mathfrak{N}}_{iR}\overline{\phi}+G_{\lambda_i}\overline{\lambda}_{iR}\overline{\phi}+G_{\chi_i}\overline{\chi}_{iR}\xi)\begin{pmatrix}\chi_i\\ \lambda_i\end{pmatrix}_L\right]+\text{H.c.}\,, \tag{29}$$

where[22]

$$\xi=\begin{pmatrix}\phi^{0\dagger}\\ \phi^-\end{pmatrix},\quad \phi=\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}, \tag{30}$$

$$G_{\alpha_a}=G_{\alpha_b}=G_{\alpha_c}\equiv G_\alpha^h\quad(\alpha=\mathcal{P},\mathfrak{N},\lambda\text{ or }\chi), \tag{31}$$

$$G_{\mathfrak{N}_a\lambda_a}=G_{\mathfrak{N}_b\lambda_b}=G_{\mathfrak{N}_c\lambda_c}\equiv\Delta_{\mathfrak{N}\lambda}^h\,, \tag{32}$$

and

$$G'_{\mathfrak{N}_a\lambda_a}=G'_{\mathfrak{N}_b\lambda_b}=G'_{\mathfrak{N}_c\lambda_c}\equiv\Delta'^h_{\mathfrak{N}\lambda}\,. \tag{33}$$

The coupling constants are chosen to be real.[23] We also choose

$$G_{\mathfrak{N}_i\lambda_i}=G'_{\mathfrak{N}_i\lambda_i} \tag{34}$$

to avoid parity-violating mass terms, and as usual

$$G_{\nu_d}=G_{\chi_d}=0 \tag{35}$$

to eliminate the appearance of right-handed neutrinos. This ensures masslessness of the neutrinos. Note that:

(1) We have allowed $(\mathfrak{N}\lambda\phi)$-coupling[24] in Eq. (29) to realize Cabibbo rotation of $\mathfrak{N}$ and λ quarks in the manner suggested in Sec. III. One may have introduced an analogous $(\mu$-$e)$ rotation in the leptonic column. This, however, does not lead to any observable consequence (see Ref. 24) if $m_{\nu_e}=m_{\nu_\mu}=0$; so we choose

$$G_{\mathfrak{N}_d\lambda_d}=G'_{\mathfrak{N}_d\lambda_d}=0\,. \tag{36}$$

(2) Subject to Eqs. (31)–(36), there are, in all, seven coupling constants to describe the masses of the fourteen fermions (not counting the neutrinos). It is easy to see that, leaving aside the hadronic $\mathfrak{N}$ and λ quarks, the masses of the others are given by

$$\begin{aligned}&m_{\mathcal{P}_a}=m_{\mathcal{P}_b}=m_{\mathcal{P}_c}=\lambda G_{\mathcal{P}}^h\,,\\ &m_{\chi_a}=m_{\chi_b}=m_{\chi_c}=\lambda G_\chi^h\,,\\ &m_e=\lambda G_{\mathfrak{N}_d}\,,\\ &m_\mu=\lambda G_{\lambda_d}\,.\end{aligned} \tag{37}$$

To find the masses of the physical $\bar{\mathfrak{N}}_i$ and $\bar{\lambda}_i$ hadronic quarks and the Cabibbo angle θ [see Eq. (10) for definition], one must diagonalize the sub-mass-matrix involving the $(\mathfrak{N}_i,\lambda_i)$ components. This gives

$$\tan2\theta=\frac{2\Delta_{\mathfrak{N}\lambda}^h}{G_\lambda^h-G_{\mathfrak{N}}^h}\,, \tag{38}$$

$$m_{\bar{\mathfrak{N}}_i}=\tfrac{1}{2}\lambda(G_{\mathfrak{N}}^h+G_\lambda^h+2\Delta_{\mathfrak{N}\lambda}^h\csc2\theta)\,, \tag{39}$$

$$m_{\bar{\lambda}_i}=\tfrac{1}{2}\lambda(G_{\mathfrak{N}}^h+G_\lambda^h-2\Delta_{\mathfrak{N}\lambda}^h\csc2\theta)\,. \tag{40}$$

Note that one must choose $G_\lambda^h\neq G_{\mathfrak{N}}^h$, otherwise $\theta=45°$.

(3) The observed approximate SU(2) symmetry of hadrons can be satisfied by choosing $G_{\mathcal{P}}^h\propto G_{\mathfrak{N}}^h>>\Delta_{\mathfrak{N}\lambda}^h$. The intrinsic mass differences between

the $\mathcal{P}$ and $\tilde{\mathfrak{N}}$ hadronic quarks, which may arise as above, provides a source of nonelectromagnetic isospin breaking.

(4) The familiar "medium strong" SU(3) breaking may be attributed simply to a $\mathfrak{N}_i$-λ_i quark mass difference arising due to spontaneous symmetry breaking. It is tempting to correlate the hadronic $\mathfrak{N}-\lambda$ mass splitting to the observed e-μ mass difference, which is on parallel footing and of similar magnitude. However, the analogy is partly spoiled by the masslessness of neutrinos, which apparently have no analogs in the hadronic columns of ψ_A.

(5) The familiar SU(3′) and chiral $SU(3')_L \times SU(3')_R$ symmetries are broken in the present model by quark mass terms (arising through spontaneous symmetry breaking) and by electromagnetic and weak interactions, but *not* by the strong interactions. In fact, it is remarkable that even after spontaneous symmetry breaking, the strong interactions in our model are invariant under the full $U(4')_L \times U(4')_R$ group (neglecting the weak mixing terms of strong and weak gauge mesons). This appears to be essentially a consequence of generating strong interactions together with weak and electromagnetic interactions by a renormalizable gauge principle. We believe this provides at least a partial answer to the often asked question: "Why do weak currents generate the symmetries of the strong interactions?" One, however, does not yet know what, if any, should set the *scale* for the breaking of various symmetries (i.e., SU(4′), SU(3′), SU(2′), chiral $U(4')_L \times U(4')_R$, $U(2')_L \times U(2')_R$, etc.), which, at present, may be introduced with arbitrary strengths through quark mass terms.

(6) Regarding SU(3″) symmetry, we have been able to arrange (because of the choice of the representations of the Higgs scalars, in particular σ_α^i and the associated vacuum expectation values) to give symmetric masses to the quarks and almost symmetric masses to the strong gauge mesons (ignoring higher-order loop corrections involving the σ's. Furthermore, the quarks do not couple[22] directly to the σ's. Thus, after the normal Higgs shifting, global SU(3″) symmetry is still well preserved in the Lagrangian except in the Higgs boson sector, in which it is lost because of the *incompleteness* of the leftover Higgs multiplet σ_α^i (in the unitary gauge). In this case, even though the Higgs bosons are coupled strongly to the strong gauge mesons, the SU(3″) breaking, which arises due to the emission and absorption of such Higgs bosons in a hadronic amplitude, is found to be only of order α compared to the symmetric amplitude (for S-matrix elements with no external Higgs bosons). This could then account for SU(3″) being an approximate observable symmetry as desired, despite the fact that the gauge symmetry is broken spontaneously. This question will be considered in more detail in a subsequent note.

V. SUMMARY

We summarize some of the peculiar features of scheme (A):

(1) The scheme generates strong interactions through an SU(3″) octet of gauge mesons, which are singlets under the familiar SU(3′) symmetry.[25] This is nonconventional in the sense that the basic strong interactions seem to owe their "existence" in such a scheme to the unfamiliar hidden SU(3″) symmetry of hadrons and *not* to the manifest SU(3′) symmetry of the low-lying world. Of course, the strong interactions thus generated respect the full $U(4')_L \times U(4')_R \times SU(3'')_{L+R}$ symmetry automatically (see Sec. IV). At the present stage, it is an assumption that the observed (strong) interactions among the known hadrons are *effective* interactions arising due to the exchange of the SU(3″) octet of gauge mesons [which we will refer to as the SU(3″) gluons]. If these gluons are not very heavy ($m \lesssim 5$ BeV, say), it should be feasible to search for their production in pairs[17] through collisions of energetic low-lying hadrons. The discovery of such gluons [which incidentally carry electric charge, even though they are SU(3′) singlets] would of course provide an interesting test of the present ideas. We should emphasize that the introduction of an SU(3″) octet of gluons to generate strong interactions instead of a singlet vector gluon (coupled to the baryonic current) is dictated by our insistence on an *anomaly-free* theory.

(2) The scheme leads to parity-violating amplitudes in hadronic processes on account of mixing of strong and weak gauge mesons. With further calculations and accumulation of experimental data, it might be possible to disentangle these contributions from that of the normal weak interactions, especially if SU(3″) is not a very strict symmetry.

(3) The main virtue of the supersymmetry group G, as presented here, has been to provide a classification scheme for the basic hadrons and leptons and thereby a possible rationale for the choice of the fundamental fermions. A special choice of G and representation F has also led to an anomaly-free gauge description of their *presently* known interactions (leaving out gravity and CP violation). On the other hand, as remarked in Sec. I, we have hardly made use of the full dynamical content of the supersymmetry group G in the sense that a whole class of gauge interactions permitted by the

model has not been considered. Most notably, the gauges which transform hadrons into leptons have been left out. Likewise, the left and right SU(4′) gauges have been treated in the most asymmetric manner. It may be of some interest to consider the implications of such gauge degrees of freedom in the model. Furthermore, by gauging the particular subgroup SU(3″) for generating strong interactions, we have built in the basic distinction between the hadrons and the leptons. A part of the future program should be to have this distinction emerge in a more natural manner from the theory.

ACKNOWLEDGMENTS

A. S. wishes to acknowledge helpful discussions with Professor J. C. Ward. J. C. P. acknowledges valuable discussions with Dr. R. N. Mohapatra, Dr. I. Bars, Professor O. W. Greenberg, Professor J. Sucher, Professor V. Teplitz, and Professor C. H. Woo. He also appreciates the hospitality of UNESCO and IAEA at the International Centre for Theoretical Physics during summer, 1972, where a large part of this work was done.

APPENDIX: GAUGE INTERACTIONS IN SCHEMES (B), (C), AND (D)

In attempting to generate a gauge theory of interactions in these three schemes [as opposed to scheme (A)], we encounter two main problems: (1) There appear, in general, baryon- and lepton-number-violating transitions, and (2) it seems difficult, if not impossible, to realize a satisfactory gauge theory of weak, electromagnetic, *and* strong interactions.

In general, (1) leads to an weak amplitude for the decay of quarks into leptons in order G_F and of the proton into leptons plus mesons in order G_F^3. In schemes (C) and (D) [but not in (B)], depending on further details, this may still be compatible with the observed lower limit on lifetime of the proton. However, in none of these schemes does there seem to be an easy solution to (2).

To illustrate these remarks, consider the generation of weak interactions in scheme (B) by assuming that the weak gauge group is SU(2). For scheme (B), the assignments are [see Eq. (9)]

$$\bar{\psi}_B = \begin{bmatrix} \mathcal{P}_a^0 & \mathcal{P}_b^+ & \mathcal{P}_c^+ & H^0 \\ \mathfrak{N}_a^- & \mathfrak{N}_b^0 & \mathfrak{N}_c^0 & H^- \\ \lambda_a^- & \lambda_b^0 & \lambda_c^0 & H'^- \\ \bar{\nu}_e & e^+ & \mu^+ & \bar{\nu}_\mu \end{bmatrix} .$$

The multiplets for the SU(2) gauge group must be chosen in accordance with the following constraints:

(1) The semileptonic decays of the known hadrons [which are SU(3″) singlets] can arise only through currents of the type $(\bar{\mathcal{P}}_i \mathfrak{N}_i)^{i=a,b,\text{or}\,c}$ involving particles in the *same* column of Ψ_B. Furthermore, to incorporate the known hadron-lepton universality in a natural manner, we ought to choose multiplets of SU(2) such that the weak current is pure SU(3″) singlet. This suggests that we introduce the multiplets

$$I = \left\{ \begin{pmatrix} \mathcal{P}_a^0 \\ \mathfrak{N}_a^- \end{pmatrix}_L , \begin{pmatrix} \mathcal{P}_b^+ \\ \mathfrak{N}_b^0 \end{pmatrix}_L , \text{ and } \begin{pmatrix} \mathcal{P}_c^+ \\ \mathfrak{N}_c^0 \end{pmatrix}_L \right\}$$

as doublets of the SU(2) gauge group.

(2) To account for the observed $|\Delta S| = 1$ semileptonic decays involving charged currents, we must assume[4] that the $(\mathfrak{N}_i, \lambda_i)_{i=a,b,c}$ quarks are related to the physical $(\bar{\mathfrak{N}}_i, \bar{\lambda}_i)_{i=a,b,c}$ quarks by Cabibbo rotation. In this case, in order to cancel the $|\Delta S| = 1$ neutral currents, we must *also* introduce the multiplets

$$J = \left\{ \begin{pmatrix} \bar{\nu}_e \\ \lambda_a^- \end{pmatrix}_L , \begin{pmatrix} e^+ \\ \lambda_b^c \end{pmatrix}_L , \begin{pmatrix} \mu^+ \\ \lambda_c^0 \end{pmatrix}_L \right\}$$

as doublets of the SU(2) group. (It is this[26] that leads to difficulties; see later.)

(3) One must of course introduce the leptonic multiplets

$$K = \left\{ \begin{pmatrix} e^+ \\ \bar{\nu}_e \end{pmatrix}_R , \begin{pmatrix} \mu^+ \\ \bar{\nu}_\mu \end{pmatrix}_R \right\}$$

as doublets of the SU(2) group to have leptonic decays.

(4) The combination $(I+J+K)$ is the *minimal set* needed to reproduce the known weak interactions without $|\Delta S| = 1$ neutral currents. The left and right components of fields not included in this set may be regarded as singlets[27] of the SU(2) group. One can introduce an additional U(1) gauge interaction to generate electromagnetism without anomalies.

It is clear that the above scheme leads to baryon- and lepton-number-violating transitions since the same gauge bosons (W^+, W^-, W^3) are exchanged between the sets I and J or J and K. Thus physical $\bar{\mathfrak{N}}_i$ and $\bar{\lambda}_i$ quarks can decay to leptons in order g^2 as follows:

$$\begin{array}{ll} & \text{Amplitude} \\ \bar{\lambda}_a^- \to \bar{\nu}_e + W^- \to \bar{\nu}_e + (e^- + \bar{\nu}_e) & (g^2 \cos\theta / m_W^2) \\ \bar{\mathfrak{N}}_a^- \to \bar{\nu}_e + W^- \to \bar{\nu}_e + (e^- + \bar{\nu}_e) & (-g^2 \sin\theta / m_W^2) \,. \end{array} \tag{A1}$$

So there would be *no stable quarks* in this scheme.

For possible decays of the observed hadrons to leptons, it is sufficient to examine the possibility

of such decays of the proton only. Assuming that the low-lying hadrons are primarily SU(3″) singlets, the proton consists of one each of (a, b, c) quarks. Consider any one of its components; for example $(\mathcal{P}_a^0 \mathcal{P}_b^+ \tilde{\mathcal{N}}_c^0)$. Assuming that the quarks and diquarks are heavier than the proton, one can easily verify that the lowest order[28] in which the proton can have a real decay is g^6; the amplitude for such a decay is proportional to $(g^6/m_W^6)\sin^3\theta$. A possible chain of virtual transitions leading to proton decay is indicated below[29]:

$$p-\begin{cases} \mathcal{P}_a^0 - \pi^+ + \tilde{\mathcal{N}}_a^- \xrightarrow{g\sin\theta} \pi^+ + (\bar{\nu}_e + W^-) \\ \qquad\qquad \xrightarrow{\ell} \pi^+ + (\bar{\nu}_e + e^- + \bar{\nu}_e)\,, \\ \mathcal{P}_b^+ - \pi^+ + \tilde{\mathcal{N}}_b^0 \xrightarrow{g\sin\theta} \pi^+ + (e^+ + W^-) \\ \qquad\qquad \xrightarrow{\ell} \pi^+ + (e^+ + e^- + \bar{\nu}_e)\,, \\ \tilde{\mathcal{N}}_c^0 \xrightarrow{g\sin\theta} \mu^+ + W^- \xrightarrow{\ell} \mu^+ + (e^- + \bar{\nu}_e)\,. \end{cases} \tag{A2}$$

A crude estimate of the lifetime of proton associated with the above decay is found to be between 10^{15} and 10^{19} sec. Even though this is of the order of or greater than the age of the universe, it is still ridiculously low compared to the experimental lower limit[30] $\tau_p > 10^{25}$ sec.

Similar considerations apply also to scheme (C); however, there is an essential difference due to the appearance of the unknown lepton E^+ in the place of μ^+, which may help remove the conflict with proton lifetime. To see this, note that if we choose the SU(2) multiplets for Ψ_c [given by Eq. (9)] analogous to the *minimal* set $(I+J+K)$ for Ψ_B, the doublet

$$\begin{pmatrix} \mu^+ \\ \lambda_c^0 \end{pmatrix}_L$$

for scheme (B) is replaced by the doublet

$$\begin{pmatrix} E^+ \\ \lambda_c^0 \end{pmatrix}_L$$

for scheme (C), the other multiplets remaining the same. On account of this, the heavy lepton E^+ would appear as an end product (instead of μ^+) in the conversion of $\tilde{\mathcal{N}}_c^0$ [see Eq. (A2)]. If $M_{E^+} > m_p$, the decay of proton involving real E^+ is forbidden. It is then easy to see that the decay of proton must at least involve the conversion of a "c" quark (say, $\tilde{\mathcal{N}}_c^0$) to an appropriate "b" or "a" quark, which in turn can decay to light leptons. As the above conversion would take place only via strong interactions and therefore would be accompanied by the emission of an appropriate strongly interacting SU(3″) octet meson, it can be shown, in this case, that the proton lifetime would be higher than the estimate given above depending upon the nature of SU(3″) symmetry.[31] For example, if SU(3″) were broken only weakly, the proton decay amplitude would involve fourth or higher powers of G_F; its lifetime may thus become long enough so as not to conflict with the experimental lower limit on τ_p. The quarks could still decay in order G_F. These remarks apply to scheme (D) as well as to (C).

We may now illustrate remark (2), mentioned in the beginning of this section. Note that the degree of freedom, which has been used to generate the known weak interactions in the above schemes, does not permit us to build strong interactions with desired symmetry and still preserve the goal of renormalizability. This is apparent simply by noting that the SU(2) gauge symmetry which treats

$$\begin{pmatrix} \bar{\nu}_e \\ \lambda_a^- \end{pmatrix}_L$$

as a doublet must be violated by strong interactions, which affect the hadron λ_a^- but not the lepton $\bar{\nu}_e$. In such a case renormalizability is lost. This argument can be further extended. In principle, a different choice of the weak gauge group [such as $O(3)$] and a cleverer choice of the multiplets may alter the situation presented here. We have not explored this possibility in detail; however, it appears difficult to change the situation substantially.

In summary, even though schemes (C) and (D) offer some intriguing possibilities which are worth noting, none of the schemes (B), (C), and (D) seems to be appropriate for realizing the goal we have sought to achieve in this paper, i.e., to generate the weak, electromagnetic, *and* strong interactions by a renormalizable gauge principle. The latter seems to be feasible only for scheme (A). We should also remind[32] the reader that the exotic possibilities pointed out in this appendix apply *exclusively* to schemes (B), (C), and (D) and not to (A), which is discussed in the text.

---

*This paper is an extended version of a preliminary draft, circulated at the Sixteenth International Conference on High Energy Physics, Batavia, Ill., 1972 (see J. D. Bjorken's review talk at the conference).

†Work supported in part by the National Science Foundation under Grant No. NSF GP 8748.

[1]See, for example, Z. Maki and T. Maskawa, Prog. Theor. Phys. 49, 1007 (1973) (earlier references

may be found in this paper). Also see A. Gamba, R. E. Marshak, and S. Okubo, Proc. Natl. Acad. Sci. USA 45, 881 (1959); S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D 2, 1285 (1970).

[2]J. Schwinger, Ann. Phys. (N.Y.) 2, 407 (1957); S. L. Glashow, Nucl. Phys. 22, 579 (1961); A. Salam and J. C. Ward, Nuovo Cimento 11, 568 (1959); A. Salam and J. C. Ward, Phys. Lett. 13, 168 (1964).

[3]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); Abdus Salam, in *Elementary Particle Theory*, edited by N. Svartholm (Almqvist, Stockholm, 1968), p. 367.

[4]S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D 2, 1285 (1970); S. Weinberg, *ibid.* 5, 1412 (1972); C. Bouchiat, J. Iliopoulos, and Ph. Meyer, Phys. Lett. 38B, 519 (1972); D. Gross and R. Jackiw, Phys. Rev. D 6, 477 (1972).

[5]M. Y. Han and Y. Nambu, Phys. Rev. 139, B1006 (1965); A. N. Tavkhelidze *et al.*, *High Energy Physics and Elementary Particles* (International Atomic Energy Agency, ICTP, Trieste, 1965), p. 763; N. Cabibbo, L. Maiani, and G. Preparata, Phys. Lett. 25B, 132 (1967).

[6]The possibility of observed leptons being composites of fractionally charged leptonic quarks is suggested by A. Salam and J. C. Pati, Phys. Lett. (to be published).

[7]S. Adler, Phys. Rev. 177, 2426 (1969); S. Okubo, *ibid.* 179, 1629 (1969).

[8]J. C. Pati and C. H. Woo, Phys. Rev. D 3, 2920 (1971).

[9]Certain variations of the Han-Nambu model suit our purpose equally well. See, for example, the version due to Cabibbo, Maiani, and Preparata (Ref. 9) and also the discussion in J. C. Pati and C. H. Woo, Phys. Rev. D 3, 1173 (1971).

[10]Note, if we do not insist on Han-Nambu structure for the hadrons, we may choose the group G to be given by SU(4) rather than SU(4′)×SU(4′), since SU(4) can accommodate three quantum numbers, which may be identified with I_3, Y, and L. In this case, if we wished to maintain a quark structure for the observed hadrons and leptons, we might consider building them "compositely" out of a single quartet of fermions $f = (\mathcal{P}, \mathfrak{N}, \lambda, l)$. However, one may verify that the $4\times4\times4$ representation will not contain the known leptons without bringing in a host of other, too-exotic subjects. The most suitable choice with G = SU(4) for the basic representation F seems to be 15. This case looks very similar to scheme (C) [see Eq. (9)], except that the baryons in the 15-plet of SU(4) would be more naturally identified with the observed (octet and singlet) baryons than with quarks.

[11]This follows by requiring that the Han-Nambu nonet of hadronic quarks carry zero lepton number. Note that other linear combinations $L = a(c' - 1.4) + b(c'' + \frac{1}{4})$ *not* included in (A)–(D) of Eq. (8) could also be considered as separate possibilities. These are, however, not distinct physically from (A)–(D). Note also that we may drop the terms $\pm\frac{1}{4}$ in Eq. (8) without altering any of the physical consequences in the present paper. These terms would be relevant only if a physical field were coupled to lepton number.

[12]For example, one can choose $B = (F - L)$ and $(F + L)$ in schemes (A) and (B), respectively, thereby assigning $B = 1$ for the hadronic quarks ($L = 0$) and $B = 0$ for the leptons. In schemes (C) and (D), some of the particles must carry *both* baryon and lepton numbers no matter how one chooses to define B in terms of F and L.

[13]We follow the notation for SU(4) generators as given by D. Amati, H. Bacry, J. Nuyts, and J. Prentki, Nuovo Cimento 34, 1732 (1964).

[14]Note that in a scheme with only one SU(4) degree of freedom, a singlet gluon gauge is the only available gauge for generating strong interactions once weak interactions have made use of $SU(2)_L$ gauges.

[15]We are indebted to Dr. R. Mohapatra for a major contribution and most helpful discussions on this section.

[16]For the present, we have not investigated whether $\lambda' \sigma^4$-like terms may be written down to yield the pattern of nonzero vacuum expectation values assumed in Eq. (22).

[17]Note that the strong gauge bosons, being primarily in the SU(3″) octet (apart from weak mixing), cannot be produced singly through collisions of normal hadrons; they have to be produced in pairs. In this case, masses around 3–5 BeV for these mesons are consistent with their absence in production experiments so far. For some discussion on production, decays, and observability of SU(3″) nonsinglet objects, see Pati and Woo (Ref. 9) and G. A. Snow, University of Paris report, 1972 (unpublished).

[18]One may also estimate σ/λ without assuming values of $m(V)$ and f separately as follows: The ratio of typical weak to strong interaction amplitude is expected to be of order

$$\frac{g^2/m_W^2}{f^2/m_V^2} \simeq \frac{\sigma^2}{\lambda^2} ,$$

which on the other hand is known to be of order 10^{-5} to 10^{-6}, consistent with Eq. (28).

[19]But for the SU(3″) suppression, one might be tempted to consider the possible relevance of such an amplitude in understanding the seemingly large discrepancy between certain recent calculations of the parity-violating amplitude in nuclear transitions and their observed values. See, for example, M. Gari, D. Dumitrescu, J. G. Zabotlisky, and H. Kummel, Phys. Lett. 35B, 19 (1971); B. Desplanques and N. Vinh Mau, *ibid.* 35B, 28 (1971).

[20]For a review of theoretical and experimental status on neutral currents see B. W. Lee, in *Proceedings of the XVI International Conference on High Energy Physics, Chicago-Batavia, Ill., 1972*, edited by J. D. Jackson and A. Roberts (NAL, Batavia, Ill. 1973), Vol. 4, p. 249.

[21]σ_α^i do not have Yukawa coupling to fermions on account of SU(3″) gauge invariance.

[22]Note that Eqs. (31)–(33) between the coupling constants are required to satisfy SU(3″) gauge invariance. This leads to SU(3″)-symmetric quark mass terms.

[23]If one allows them to be complex, one will obtain parity- and CP-violating mass terms. Such possibilities could be relevant for inducing CP violation.

[24]Note that in general a $(\mathcal{P}\chi\phi)$ coupling is also allowed, which would lead to a rotation of the $(\mathcal{P}, \chi)$ fields. The net effect of simultaneous rotations of the $(\mathfrak{N}, \lambda)$ and the $(\mathcal{P}, \chi)$ quarks by angles θ_1 and θ_2, respectively, is that (for processes of interest) the Cabibbo angle is $(\theta_1 - \theta_2)$. We do not introduce $(\mathcal{P}, \chi)$ rotation. *Note added in proof.* A related remark is that if $m_{\nu_e} = m_{\nu_\mu}$ and there is no nondiagonal mass term connecting $\nu_e \longleftrightarrow \nu_\mu$, one can

always rotate the two neutrinos so as to counteract the effect of $(\mu\text{-}e)$ rotation. The $\mu \to e + \gamma$- decay experiment and the famous two neutrino experiments are, therefore, insensitive to such rotations. We thank Dr. W. Alles for this last remark.

[25]Note the distinction in this case from the models of A. Salam and J. Strathdee [Nuovo Cimento 11A, 397 (1972) (in particular p. 426)] and I. Bars, M. B. Halpern, and M. Yoshimura [Phys. Rev. Lett. 29, 969 (1972)], who also propose a gauge theory of the strong, electromagnetic, and weak interactions. In these models, hadrons are *not* coupled directly to the weak gauge bosons, and their strong interactions are mediated by a nonet of vector and a nonet of axial-vector gauge mesons, which transform as (8 + 1) under familiar SU(3). Note that, for low-lying hadrons, in our scheme, SU(3′) and SU(3) are synonymous.

[26]Note that the leptons in set J cannot be replaced by the hadrons (H^0, H^-, H'^-) because of their charges.

[27]On the other hand, for the sake of symmetry, one might have introduced an extended set by treating the four columns of Ψ_B in an identical manner and likewise the four rows. This extended set will be

$$(I+J+K)+\left\{\begin{pmatrix}H^0\\H^-\end{pmatrix}_L,\begin{pmatrix}\nu_\mu\\H'^-\end{pmatrix}_L,\begin{pmatrix}\lambda_b^0\\\lambda_a^-\end{pmatrix}_R,\begin{pmatrix}\lambda_c^0\\H'^-\end{pmatrix}_R,\right.$$
$$\left.\begin{pmatrix}\mathfrak{N}_b^0\\\mathfrak{N}_a^-\end{pmatrix}_R,\begin{pmatrix}\mathfrak{N}_c^0\\H^-\end{pmatrix}_R,\begin{pmatrix}\varphi_b^+\\\varphi_a^0\end{pmatrix}_R,\begin{pmatrix}\varphi_c^+\\H^0\end{pmatrix}_R\right\}.$$

For scheme (B), the consequences (discussed in the text) are the same for the extended set as it is for the minimal set. [This is not so for schemes (C) and (D).]

[28]Note that the loop diagram of order g^4 cannot contribute in this case, since it cannot get rid of quarks entirely.

[29]Note that the emission of two π^+ mesons is essential, since it allows the conversion of φ_a^0 and φ_b^+ to the physical $\tilde{\mathfrak{N}}_a^-$ and $\tilde{\mathfrak{N}}_b^0$ quarks, respectively. The latter contain λ_a^- and λ_b^0 by an amount given by $\sin\theta$; it is the λ quarks that can be converted to (lepton + W).

[30]Particle Data Group, Phys. Lett. 39B, 1 (1972) and specific references therein.

[31]If SU(3′′) were an exact symmetry [in Scheme (C)], one could realize the intriguing possibility that the proton is absolutely stable, even if the quarks are not. In this case, at least one other particle (in addition to the proton) must also be absolutely stable. This may be the E^+ or a SU(3′′) nonsinglet hadron (distinct from quarks).

[32]As a further side remark, it is worth noting that in each of these schemes [(A) through (D)] one may introduce a singlet gauge particle coupled to the fermion number, which could be absolutely conserved. By demanding that the corresponding interaction be long-range (and very very weak), one could, for example, avoid the transcendence of fermion number in black-hole physics, though baryon and lepton numbers may still be transcended individually.

VOLUME 31, NUMBER 10 PHYSICAL REVIEW LETTERS 3 SEPTEMBER 1973

Is Baryon Number Conserved?

Jogesh C. Pati*
International Centre for Theoretical Physics, Trieste, Italy, and Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

and

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England
(Received 3 August 1973)

We suggest that baryon-number conservation may not be absolute and that an integrally charged quark may disintegrate into two leptons and an antilepton with a coupling strength $G_B m_p^2 \lesssim 10^{-9}$. On the other hand, if quarks are much heavier than low-lying hadrons, the decay of a three-quark system like the proton is highly forbidden (proton lifetime $\gtrsim 10^{28}$ y). Motivation for these ideas appears to arise within a unified theory of hadrons and leptons and their gauge interactions. We emphasize the consequences of such a possibility for real quark searches.

It is part of general belief in particle physics that conservation of baryon number is an absolute law of nature. Such a notion is but natural when one considers the extraordinary stability of the lightest known baryon, the proton, with a lifetime in excess of 10^{35} sec. In this note we wish to question whether this apparent proton stability truly reflects the conservation of baryon number to a similar degree.

Specifically, we have in mind the following possibility. Assume that the proton is made up in some sense of three quarks, each quark (q) carrying baryon number $B=1$ and an integral electric charge. Assume that quarks and diquarks (if the latter exist) are heavier than the low-lying hadrons. Assume further that a quark can decay into two of the known leptons ($l=\nu_e, e^-, \nu_\mu, \mu^-$) and an antilepton, the decay being described by an effective Lagrangian

$$\mathcal{L}_{\text{eff}}=(G_B/\sqrt{2})(\bar{q}l)(\bar{l}\bar{l})+\text{H.c.} \qquad (1)$$

This decay violates conservation of baryon and lepton (L) numbers, but conserves fermion number F where $F=B+L$.

Our point is this: If states with quark and diquark quantum numbers are heavier than the low-lying hadrons, a proton ($B=3$, $F=3$) can have a real decay only to three leptons plus mesons, or four leptons plus an antilepton, etc. Since this involves a violation of baryon number by three units, the lowest-order amplitude in G_B in which a proton can decay[1] is therefore G_B^3. Assume that $G_B m_p^2 \lesssim 10^{-9}$. (We give later our theoretical reasons for G_B being less than the decay constant for $|\Delta S|\neq 0$ neutral semileptonic transitions, which is of order $G_F\alpha^2$ empirically.) We then find that the decay $q \to l+l+\bar{l}$ may be associated with a lifetime as short as 10^{-10} sec, say (depending on quark mass), whereas the proton's lifetime could still be far in excess of 10^{35} sec because of the high degree of forbiddenness of its decay. Such a model would therefore show that (i) quarks (if they exist) may exhibit unexpected decay properties involving violation of baryon number as well as lepton number without conflicting with the observed degree of stability of the proton, and (ii) there is the possibility that there is no stable quark, contrary to present belief. This may be one reason why conventional searches for quarks have been unsuccessful (especially if $\tau_q \lesssim 10^{-10}$ sec). (Note that if quarks were fractionally charged, electric-charge con-

servation would imply that a stable quark must exist unless there were lighter fractionally charged leptons[2] into which the quarks could decay.)

Before giving the motivation which led us to consider an effective interaction of the type (1) —*and we stress that the general considerations above hold irrespective of any specific model* —we give some order-of-magnitude estimates for typical quark and proton decay widths (see Table I). The quark decay width follows directly from Eq. (1):

$$\Gamma(q\to l+l+\bar{l}) = G_B{}^2 m_q{}^5/24(2\pi)^3 = \lambda^2(3\times10^7\ \text{sec}^{-1})\left(\frac{m_q}{10\ \text{GeV}}\right)^5, \quad (2)$$

where we have put $G_B=\lambda G_p\alpha^2$ (we estimate $\lambda\lesssim 1$; see below). Two typical proton decay amplitudes (M) and corresponding rates ($\Gamma=|M|^2\rho$, where ρ is the phase-space factor) are listed in Table I. The quantities m_q and m_p are quark[3] and proton masses; A_4 and A_5 are numerical factors arising out of four- and five-particle phase-space integrations, which depend somewhat on precise matrix elements [usually they are $\gg 1$; for example $A_5=12$, see Eq. (2)]; Λ is a cutoff, which may be of the order of a few GeV (as in the K_L-K_S mass-difference calculation). Using Table I, we obtain for a typical proton decay[4]

$$\Gamma(p\to 3l+\pi) = 2\times10^{-39}\lambda^6/A_4\ \text{sec}^{-1} \quad (3)$$

for $m_q=10$ GeV and $\Lambda=2m_p$. Thus for $\lambda\lesssim 1$, $\Gamma_p \lesssim 2\times10^{-39}\ \text{sec}^{-1}$. Although one may not take the precise estimates given above too seriously, the main point we wish to emphasize is that the high degree of forbiddenness of the proton decay relative to the quark decay appears to be sufficient to lay open the possibility that the proton may be comfortably stable ($\tau_p>10^{28}$ yr) and yet the quarks decaying into leptons sufficiently short lived.

Since $G_B m_p{}^2\approx 10^{-9}$ implies a characteristic energy (or intermediate meson mass) of the order of 3×10^4 GeV, one may expect that at (cosmic-ray) energies of this order, reaction rates for the processes $e+p$ or $p+p\to$ leptons + antileptons, etc., would attain unitarity limit and effectively become strong. Thus a study of multilepton-induced showers at high cosmic-ray energies may be one way of testing the ideas presented above. A clearer test could hopefully be provided by extending the search for real integrally charged quarks at the CERN intersecting storage rings and the National Accelerator Laboratory to detect possible disintegration of quarks into *energetic* leptons ($q\to l+l+\bar{l}$, $q\to l+\pi$, $q\to l+\gamma$, ..., etc.). One should allow for $\tau_q\gtrsim 10^{-10}$ sec on the one hand and perhaps[5] as short as 10^{-12} to 10^{-13} sec on the other.

Our basic motivation for B nonconservation comes from a recent attempt[1,6] at a gauge theory of strong, weak, and electromagnetic interactions. To construct a unified, anomaly-free, renormalizable gauge model we suggested that a system of twelve integrally charged quarks (nine of Han-Nambu variety, and three charmed quarks) plus the four known leptons ($\nu_e, e^-, \mu^-, \nu_\mu$) be combined in a ($\underline{4}, \underline{4}^*$) representation F of an $SU(4)_{L+R}{}' \otimes SU(4)_{L+R}{}''$ group structure,

$$F=\begin{pmatrix} \mathcal{P}_a{}^0 & \mathcal{P}_b{}^+ & \mathcal{P}_c{}^+ & \nu_e \\ \mathcal{N}_a{}^- & \mathcal{N}_b{}^0 & \mathcal{N}_c{}^0 & e^- \\ \Lambda_a{}^- & \Lambda_b{}^0 & \Lambda_c{}^0 & \mu^- \\ \chi_a{}^0 & \chi_b{}^+ & \chi_c{}^+ & \nu_\mu \end{pmatrix}. \quad (4)$$

The strong interactions were introduced by gauging an $SU(3)_{L+R}{}''$ subgroup of $SU(4)_{L+R}{}''$, the conventional weak interactions by gauging an $[SU_L(2)']$ subgroup of $SU(4)_L{}'$, while the electromagnetic gauges spanned over generators of *both* $SU(4)'$ and $SU(4)''$, i.e.,

$$Q=(I_3{}'+\tfrac{1}{2}Y'-\tfrac{2}{3}C')_{L+R} +(I_3{}''+\tfrac{1}{2}Y''-\tfrac{2}{3}C'')_{L+R}. \quad (5)$$

TABLE I. Estimates of typical proton decay modes. We have not exhibited factors of $(2\pi)^{-n}$ $(n>0)$ in the matrix element M, which usually arise from virtual loops. These suppress proton decay rate still further.

Decay	M	ρ	Γ
$p\to l+l+\bar{l}+\pi$	$\left(\frac{G_B}{\sqrt{2}}\right)^3\left(\frac{\Lambda}{m_q}\right)^3\Lambda^3$	$\frac{M_p{}^7}{A_4(2\pi)^5}$	$m_p\left(\frac{G_B m_p{}^2}{\sqrt{2}}\right)^6\left(\frac{\Lambda}{m_q}\right)^6\left(\frac{\Lambda}{m_p}\right)^6\frac{1}{A_4(2\pi)^5}$
$p\to l+l+l+\bar{l}$	$\left(\frac{G_B}{\sqrt{2}}\right)^3\left(\frac{\Lambda}{m_q}\right)^3\Lambda$	$\frac{M_p{}^{11}}{A_5(2\pi)^7}$	$m_p\left(\frac{G_B m_p{}^2}{\sqrt{2}}\right)^6\left(\frac{\Lambda}{m_q}\right)^6\left(\frac{\Lambda}{m_p}\right)\frac{1}{A_5(2\pi)^7}$

VOLUME 31, NUMBER 10 PHYSICAL REVIEW LETTERS 3 SEPTEMBER 1973

The theory at this stage had no exotic consequences except for the unusual unification of hadronic matter ($B=1$, $L=0$) with leptonic matter ($B=0$, $L=1$) within the *same* multiplet of a common symmetry structure $SU(4)'\otimes SU(4)''$.

But in order that such a unification be dynamically compelling, one must gauge sufficient degrees of freedom (consistent with established[7] selection rules) to ensure transformability of leptons into baryons. This still does not imply nonconservation of baryon-lepton numbers because appropriate gauge bosons could carry these numbers. What we want to show is that if in addition to the subgroups mentioned above we had also gauged the remaining degrees of freedom of $SU(4)'$ and $SU(4)''$, or even a non-Abelian subset (stated below) such that the electric current is expressed as a sum of *non-Abelian* currents from both groups $SU(4)'$ and $SU(4)''$, the requirement of electric-charge conservation—expressed in terms of masslessness of the photon—together with the twin requirements of renormalizability and appropriate[8] massiveness of all other gauge bosons, necessarily appears to lead to lepton-baryon-number nonconservation.

To demonstrate this, consider the local gauge structure[9,10] $SU(4)_L'\otimes SU(4)_R'\otimes SU(4)_{L+R}''$ {although the essential ingredients of the argument become manifest already at the stage of the smaller gauge symmetry $[SU(2)_L']\otimes[SU(2)_R']\otimes SU(4)_{L+R}''$, preferable for reasons connected with anomalies}. Let $W_{ij}^{L,R}$ and V_{ij} ($i,j=1,2,3,4$) represent the fifteen-plet of gauge mesons associated with the groups $SU(4)_{L,R}'$ and $SU(4)_{L+R}''$ with $J_{ij}'^{L,R}$ and J_{ij}'' denoting the associated currents. Note that the quantum numbers B, L associated with these currents are as follows:

$J_{ij}'^{L,R}$ (all i and j), $B=L=0$;

J_{ij}'' ($i,j=1,2,3$; $i=j=4$), $B=L=0$;

J_{ij}'' ($i=1,2,3$; $j=4$), $B=1$, $L=-1$;

J_{ij}'' ($i=4$; $j=1,2,3$), $B=-1$, $L=+1$;

the last two groups being exotic. The point to be emphasized is that unless the theory forces a mixing of the nonexotic currents ($B=L=0$) with the exotic ones (with $B\neq 0$, $L\neq 0$), the mere existence of such currents and the corresponding gauge mesons would not violate B-L conservation. Such a mixing, however, appears necessary if one attempts to give masses to all gauge mesons (with the sole exception of the photon) through a Higgs-Kibble mechanism. This is because with the electric charge expressed as a sum of non-Abelian generators [as given by Eq. (5)], appropriate massiveness of all gauge bosons other than the photon can be secured only[11] by postulating, among other representations, the existence of a mixed representation of Higgs-Kibble σ particles—typically a $(\underline{1},\underline{4},\underline{4}^*)$ representation of $SU(4)_L'\otimes SU(4)_R'\otimes SU(\underline{4})_{L+R}''$—with expectation values in the sequence indicated:

$$\langle\sigma\rangle_{ij}=\alpha_i\delta_{ij}\quad(i,j=1,2,3,4). \tag{6}$$

Quite clearly, the gauge term in the Lagrangian $|g_R W^R\langle\sigma\rangle+f\langle\sigma\rangle V|^2$ [where g_R and f are the coupling constants associated with weak $SU(4)_R'$ and strong $SU(4)_{L+R}'$ gauge groups, respectively] induces not only the appropriate mixing of neutral V's and the W's which go to make up the photon but also a mixing of the exotic V's with W^R's, coupled to $B=0=L$ currents. It is this mixing which is responsible for B-L nonconservation.

A typical term in the effective current-current Lagrangian induced by the above mixing is of the form

$$\mathcal{L}_B=G_B(\bar{\mathcal{O}}_e{}^0\nu_e+\bar{\mathcal{N}}_e{}^-e^-+\bar{\lambda}_e{}^-\mu^-+\bar{\chi}^0\nu_\mu)\times(C^0+\bar{\nu}_\mu\nu_e)_R+\text{H.c.}, \tag{7}$$

where $C^0=(\bar{\mathcal{O}}_e{}^0\chi_e{}^0+\bar{\mathcal{O}}_b\chi_b{}^++\bar{\mathcal{O}}_c\chi_c{}^+)_R$ is the charm current and $\mathcal{L}_B$ is part of the structure

$$\alpha_1\alpha_4J_{14}''J_{41}'^R+\alpha_2\alpha_4J_{24}''J_{42}'^R+\alpha_3\alpha_4J_{34}''J_{43}'^R+\text{H.c.} \tag{8}$$

To obtain an estimate of G_B, first note that exchanges of the exotic V mesons [as well as exchanges of W_{3i}^R ($i=1,2$)] induce neutral semileptonic $|\Delta S|\neq 0$ transitions with effective strength $\simeq f^2/m_x^2$, where m_x is an exotic meson mass. In order that this be consistent with observed limits, f^2/m_x^2 must be $\lesssim G_F\alpha^2$. Thus[12] $m_x\gtrsim f(3\times10^4$ GeV). Since $G_B=\kappa(f^2/m_x^2)$, where κ is a mixing parameter (whose detailed value depends on the mass matrix), we infer that empirically $G_B\lesssim G_F\alpha^2$.

It is amusing to note that, depending on the details of the model chosen for the Higgs-Kibble scalars (and whether μ^- or e^- is the "strange" lepton), one will encounter varying selection rules for quark and proton decays. The precise structure of the B-L-nonconserving interaction obtained above leads to quark decays of the fol-

lowing variety:

(A) $(\mathcal{P}_a{}^0, \mathcal{R}_a{}^-, \lambda_a{}^-) \to (\nu_e, e^-, \mu^-) + (\nu_\mu + \bar{\nu}_e)$.

(B) $(\mathcal{P}_b{}^+, \mathcal{R}_b{}^0, \lambda_b{}^0) \to (\nu_e, e^-, \mu^-) + (\nu_\mu + e^+)$.

(C) $(\mathcal{P}_c{}^+, \mathcal{R}_c{}^0, \lambda_c{}^0) \to (\nu_e, e^-, \mu^-) + (\nu_\mu + \mu^+)$.

These would lead to a proton's decay to seven or nine leptons (including antileptons), but not to $3l + \pi$ or $4l + \bar{l}$. One may note that within the smaller gauge structure $\{[SU(2)_L{}'] \otimes [SU(2)_R{}'] \otimes SU(4)_{L+R}{}''\}$, $B-L$ nonconservation proceeds only through the term $J_{34}{}''(J_{43}{}'^R + J_{12}{}'^R)$. This will allow decays of the type (C), but not of (A) and (B). In this case, since the proton is made up of (a, b, c) quarks, one can show that its decay is further suppressed[13] at least by additional factors of α.

Since the characteristic energies ($\approx 10^4$–10^5 GeV) discussed above (which, we stress, represent a *new scale* in particle physics) are not the energies encountered in normal star interiors, one does not expect significant astrophysical implications of $B-L$ nonconservation except in the early stages of the universe (when baryons may have been produced from energetic lepton-lepton collisions or vice versa) and possibly in black-holes and quasars.

To conclude, while arguments based on a particular set of theoretical ideas are never compelling, the general considerations on forbiddenness of proton decay in a heavy quark model remain and need experimental verification. If the gauge ideas are correct, we find it amusing that the only known massless gauge particle is the photon. Could it be that the electric charge is the only non-Abelian[14] conserved charge in nature?

We thank Professor P. Budini, Professor R. Dalitz, Professor R. Glasser, Professor M. Goldhaber, Professor D. Sciama, Professor A. Sirlin, and Dr. J. Strathdee for discussions. J.C.P. wishes to acknowledge the hospitality of the International Atomic Energy Agency and UNESCO during his stay at the International Centre for Theoretical Physics, Trieste, in the summer of 1973.

---

*Work supported in part by the National Science Foundation under Grant No. NSF GP 20709.

[1]This remark orginates from Appendix A of J. C. Pati and A. Salam, Phys. Rev. D 8, 1240 (1973). The motivation for baryon-lepton-number nonconservation in the present paper is very different from that presented in this Appendix.

[2]A. Salam and J. C. Pati, Phys. Lett. 43B, 311 (1973).

[3]The characteristic appearance of $(G_B/m_q)^3$ in the proton-decay matrix elements is due to the fact that the unitarity sum for ImM necessarily contains a product of the three physical-quark-decay matrix elements and hence the real part is proportional to $\sum_E G_B{}^3 (m_q - E)^{-3}$.

[4]Note that of the value given in Eq. (3), 6×10^{-35} sec^{-1} comes from a characteristic three-body decay width alone (if such a decay were allowed) i.e., $\Gamma(p \to 3l) = (96\pi^3)^{-1}(2^{-3/2} G_B{}^3 m_p{}^6)^2 m_p$, with $G_B m_p{}^2 = G_F \alpha^2$.

[5]Because of the uncertainty in the estimate of the proton-decay matrix element, it is possible that the quark lifetime could even be of order 10^{-12} to 10^{-13} sec without conflicting with the proton lifetime, especially if additional selection rules are involved in proton decay (see also remarks below).

[6]C. Itoh, T. Miamikawa, K. Miura, and T. Watanabe, "Unified Gauge Theory of Weak Electromagnetic and Strong Interactions" (to be published). This model is similar to that of Ref. 1, except that quarks are fractionally charged while leptons are integrally charged and the gauge bosons are massless. There would be no possibility of quarks decaying into leptons in this scheme.

[7]The whole purpose of the introductory section was to question whether baryon conservation is indeed all that well established.

[8]Consistent with approximate global SU(3)'' symmetry and effectively weak lepton interactions [so that $\alpha_1 \simeq \alpha_2 \simeq \alpha_3 \neq 0$ and $\alpha_4 \neq 0$ and large in Eq. (6)].

[9]The desirability of gauging an extended group structure was suggested in Ref. 1. The bigger group structure $SU(4)_L{}' \otimes SU(4)_R{}' \otimes SU(4)_{L+R}{}''$ leads to anomalies. On the other hand a simple and elegant scheme is obtained within the smaller gauge symmetry $SU(2)_L{}' \otimes SU(2)_R{}' \otimes SU(4)_{L+R}{}''$, which we consider in some detail in a forthcoming note.

[10]D. Ross, to be published, has independently considered the consequences of gauging SU(4)' × SU(4)'' within the unified model of Ref. 1. His work confirms the conclusion regarding $B-L$ nonconservation in such a scheme.

[11]The necessity for such a representation involves a longer discussion and will be given elsewhere.

[12]It is well known that suppression due to large masses is not retained in general by loop diagrams (see for example, S. Weinberg, Phys. Rev., to be published. Preliminary studies reveal that in the model presented here B and L nonconservations are suppressed by heavy masses even in loop diagrams. This is to be considered in detail elsewhere.

[13]It is even possible that the proton could be made absolutely stable in this model, provided there exists an additional particle (meson) in the theory, which is also absolutely stable and heavier than the proton (see Appendix A of Ref. 1 for details of this mechanism).

[14]There is, of course, still the possibility of gauging the U(1) Abelian generator, corresponding to fermion-number conservation in the theory.

PHYSICAL REVIEW D VOLUME 10, NUMBER 1 1 JULY 1974

Lepton number as the fourth "color"

Jogesh C. Pati*
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England
(Received 25 February 1974)

Universal strong, weak, and electromagnetic interactions of leptons and hadrons are generated by gauging a non-Abelian renormalizable anomaly-free subgroup of the fundamental symmetry structure $SU(4)_L \times SU(4)_R \times SU(4')$, which unites three quartets of "colored" baryonic quarks and the quartet of known leptons into 16-folds of chiral fermionic multiplets, with lepton number treated as the fourth "color" quantum number. Experimental consequences of this scheme are discussed. These include (1) the emergence and effects of exotic gauge mesons carrying both baryonic as well as leptonic quantum numbers, particularly in semileptonic processes, (2) the manifestation of anomalous strong interactions among leptonic and semileptonic processes at high energies, (3) the independent possibility of baryon-lepton number violation in quark and proton decays, and (4) the occurrence of $(V+A)$ weak-current effects.

I. INTRODUCTION

In two recent notes[1,2] we proposed grouping baryonic quarks ($B=1$) and leptons ($L=1$) together as members of the same fermionic multiplet ($F=B+L=1$) and generating weak and electromagnetic as well as strong interactions through a gauging of the symmetry group of this multiplet. In the first place, this postulate of a common fermionic multiplet for all fundamental matter guarantees that in any model of weak interactions, the same[3] $[\frac{1}{2}(1+\gamma_5)]$ helicity projection manifests itself for leptons (as contrasted to antileptons) as is manifested for baryonic quarks. In the second place, the gauging of the symmetry group of matter ensures that all interactions, weak and electromagnetic as well as strong, are universal with respect to baryons and leptons. While the detailed dynamical model of gauge interactions clearly depends on the precise symmetry group one may choose for the fermions (quarks + leptons), it must be emphasized that all such models share the following three characteristics:

(1) Among the gauge particles, there must exist *exotic* particles (X particles) carrying both baryonic as well as leptonic quantum numbers. In the lowest orders of perturbation theory such particles would mediate semileptonic interactions only.

(2) If all allowed gauge degrees of freedom are realized through appropriate gauge bosons, the universality of gauge interactions implies that leptonic and semileptonic interactions must eventually become strong. The asymmetric response of leptons and baryons to strong interactions at presently attained energies would then be interpreted as a "low"-energy phenomenon.[4]

(3) If appropriate spontaneous symmetry-breaking is postulated, there is the (logically independent) possibility of baryonic quarks transforming into leptons, with a violation of baryon and lepton number conservation (though the fermion number $F=B+L$ is still conserved).

In this paper, we wish to concentrate on one class of fermionic models for quarks and leptons. This class was briefly motivated in II; here we shall be concerned with the experimental consequences. However, we wish to emphasize once again that the notion that all fundamental matter is of one variety and that this lepton-baryon unification leads to the three general consequences enumerated above is something which lies at a level much deeper than the particular models discussed in this paper, which may or may not need modifications as new experimental facts emerge, and it is this unification which we principally wish to stress.

II. THE "BASIC" MODEL AND ITS VARIANTS

The central assumption of the "basic" model we propose is that quarks carry four "colors:" Three of these (a, b, and c in our notation; red, blue, and white in the more familiar terminology) represent baryonic matter ($B=1$), and the *fourth* (d or lilac) represents lepton number L.[5] The unification of baryonic and leptonic matter arises by extending the gauge symmetry $SU(3')$ of the three colors[1] (a, b, c) to $SU(4')$ of the four colors (a, b, c, d). We shall assume that the fifteen (1^-) gauge mesons corresponding to $SU(4')$ generate

strong interactions with $f^2/4\pi \simeq 1-10$.

Accepting that (spin-$\frac{1}{2}$) quarks form quartets with four *valency* quantum numbers ($I_3 = \pm\frac{1}{2}$, strangeness S, and charm C), with an underlying group structure $SU(4)_L \times SU(4)_R$, the full *global* structure we are postulating [and one which contains the classification symmetry[6] $SU(3)\times SU(3')$ of hadrons] corresponds to

$$G = SU(4)_L \times SU(4)_R \times SU(4') . \tag{1}$$

This symmetry is mathematically realized by a composite structure,[7]

$$\psi_{L,R} = \begin{bmatrix} \mathcal{P} \\ \mathfrak{N} \\ \lambda \\ \chi \end{bmatrix}_{L,R} \otimes (a, b, c, d) ,$$

where the (spin-$\frac{1}{2}$) column $(\mathcal{P}, \mathfrak{N}, \lambda, \chi)$ indicates valency and the (spin-zero) row (a, b, c, d) indicates color degrees of freedom. A *physical* realization of this structure is provided by the following two 16-fold fermions.

A. Fermions

$$\Psi_{L,R} = \begin{bmatrix} \mathcal{P}_a & \mathcal{P}_b & \mathcal{P}_c & \mathcal{P}_d = \nu \\ \mathfrak{N}_a & \mathfrak{N}_b & \mathfrak{N}_c & \mathfrak{N}_d = e^- \\ \lambda_a & \lambda_b & \lambda_c & \lambda_d = \mu^- \\ \chi_a & \chi_b & \chi_c & \chi_d = \nu' \end{bmatrix}_{L,R} = [\Psi^i_\alpha]_{L,R} .$$

Their transformation properties are

$$\Psi_L = (4, 1, \bar{4})_G ,$$

$$\Psi_R = (1, 4, \bar{4})_G ,$$

$$i = \mathcal{P}, \mathfrak{N}, \lambda, \chi ,$$

$$\alpha = a, b, c, d .$$

These multiplets contain twelve baryonic quarks, together with a lepton quartet which we have identified with the *known* leptons.[8]

B. Gauge mesons in the "basic" model

The maximal[9] anomaly-free (renormalizable) subgroup of the valency group $SU(4)_L \times SU(4)_R$ is $Sp(4)_L \times Sp(4)_R$, for which each quark (or lepton) quartet transforms as a 4-component internal-symmetry spinor. Likewise (without a doubling of quarks and leptons), the maximal anomaly-free strong gauge group, which contains the strong $SU(3')$ as a subgroup, is $SU(4')_{L+R}$. Accepting the principle that a symmetry group is manifested only through the dynamical interactions of the theory, we should gauge

$$Sp(4)_L \times Sp(4)_R \times SU(4')_{L+R} ,$$

yielding a total of $10+10+15=35$ gauge fields. However, most of the essential features of the model are retained, insofar as its physical predictions are concerned, if we simplify our considerations and choose to work with the smaller local subgroup

$$\mathcal{G} = [SU(2)_L^{I+II}] \times [SU(2)_R^{I+II}] \times SU(4')_{L+R} , \tag{2}$$

for which Ψ_L and Ψ_R transform as $(2+2, 1, \bar{4})$ and $(1, 2+2, \bar{4})$, respectively. In the sequel we shall do this. The groups $SU(2)_L^I$ and $SU(2)_L^{II}$ act on the $(\mathcal{P}, \mathfrak{N})_L$ and $(\lambda, \chi)_L$ indices, respectively (or rather, on the corresponding Cabibbo-rotated fields; see Sec. IV), while $SU(2)_L^{I+II}$ is their diagonal sum.

Before we discuss the structure of the local gauges, let us list some of the general features of the proposed gauge scheme:

(1) In contrast to the scheme proposed in I [where only the subgroup $SU(3')\times U(1')$ of $SU(4')$ was gauged], the present scheme treats leptons and baryons universally even so far as strong gauge couplings are concerned. As will be seen in Sec. III, the presently observed differences between leptons and baryons in this regard will be attributed (through a mechanism of spontaneous symmetry breaking) to a heavy mass of those strong gauge mesons, which interact with the leptons. The advantages[10] of the restricted gauge scheme proposed in I, in respect of effective strong interactions generated through the mediation of a relatively light $SU(3')$ octet, are of course preserved in the present scheme.

(2) If the bare coupling constants $g_L^{(0)}$ and $g_R^{(0)}$ are equal, the Lagrangian would exhibit *complete symmetry* between left and right helicities insofar as fermion–gauge-meson interactions are concerned. The observed left-right asymmetry (i.e., parity violation) at low and medium energies may thus be ascribed to heavier masses of the "right-hand" weak gauge mesons compared to the "left-hand" ones, introduced via spontaneous symmetry breaking.

(3) An advantage of gauging the full $SU(4')$ *and* the right-hand gauges[11] (in contrast to the restricted scheme of I) is that it is possible to generate electromagnetism without ever introducing an Abelian U(1)-gauge group for this purpose. The elimination of an Abelian quantum-number contribution to electric charge is most desirable in understanding why electric charge is so quantized. Furthermore, the absence of U(1) may have importance in securing "asymptotic freedom" for the complete theory, including electromagnetism.

Below we list the set of 21 $(=3+3+15)$ gauge particles corresponding to the "basic" model with the local gauge group $\mathcal{G}$. These are

$$W_L = (\underline{3},1,1) = \begin{bmatrix} \tau\cdot W_L & 0 \\ 0 & \tau_1(\tau\cdot W_L)\tau_1 \end{bmatrix}, \qquad \text{coupling } \frac{g_L^2}{4\pi} \simeq \alpha$$

$$W_R = (1,\underline{3},1) = \begin{bmatrix} \tau\cdot W_R & 0 \\ 0 & \tau_1(\tau\cdot W_R)\tau_1 \end{bmatrix}, \qquad \text{coupling } \frac{g_R^2}{4\pi} \simeq \alpha \tag{3}$$

$$V = (1,1,\underline{15}) = \frac{1}{\sqrt{2}} \begin{bmatrix} V(8) - \dfrac{S^0\times 1}{\sqrt{12}} & X \\ \overline{X} & (\frac{3}{4})^{1/2}S^0 \end{bmatrix}, \qquad \text{coupling } \frac{f^2}{4\pi} \simeq 1\sim 10\,.$$

$V(8)$ in the 3×3 matrix block for V denotes the SU(3′) color octet of gauge mesons consisting of V_ρ, V_{K^*}, and V_ϑ. X is an exotic ($B=+1$, $L=-1$) SU(3′) triplet[12] with members $(X^0, X^-, X^{-\prime})$, and S^0 is an SU(3′) singlet.

Defining $\nabla_\mu\Psi = \partial_\mu\Psi + igW\Psi_\mu - if\Psi V_\mu$, the Fermi Lagrangian is given by

$$-\mathrm{Tr}[\overline{\Psi}_L(\gamma_\mu\nabla_\mu)_L\Psi_L + (L\rightarrow R)]\,. \tag{4}$$

Thus

$$L_{\text{int}} = g_L \sum_{\alpha=a,b,c,d} (\overline{\mathcal{P}}_\alpha\overline{\mathcal{N}}_\alpha\overline{\lambda}_\alpha\overline{\chi}_\alpha)_L (W_L)_\mu\gamma_\mu[\tfrac{1}{2}(1+\gamma_5)] \begin{bmatrix} \mathcal{P}_\alpha \\ \mathcal{N}_\alpha \\ \lambda_\alpha \\ \chi_\alpha \end{bmatrix}_L + g_R \sum_{\alpha=a,b,c,d} (\overline{\mathcal{P}}_\alpha\overline{\mathcal{N}}_\alpha\overline{\lambda}_\alpha\overline{\chi}_\alpha)_R (W_R)_\mu\gamma_\mu[\tfrac{1}{2}(1-\gamma_5)] \begin{bmatrix} \mathcal{P}_\alpha \\ \mathcal{N}_\alpha \\ \lambda_\alpha \\ \chi_\alpha \end{bmatrix}_R$$

$$+ f \sum_{i=\mathcal{P},\mathcal{N},\lambda,\chi} (\overline{\psi}_a^i\overline{\psi}_b^i\overline{\psi}_c^i\overline{\psi}_d^i)_{L+R} V_\mu\gamma_\mu \begin{bmatrix} \psi_a^i \\ \psi_b^i \\ \psi_c^i \\ \psi_d^i \end{bmatrix}_{L+R}. \tag{5}$$

The complete Lagrangian (after Cabibbo rotations) is exhibited in (16a) and (16c) of Secs. IV B and IV C.

C. The photon

To identify the photon field we must fix on a charge formula for the fermionic multiplet. It is easy to show that the postulate that the known baryons are three baryonic quark composites and have $F_3' = F_8' = 0$ leaves us with just the following choice for the charge operator[13] Q:

$$Q = I_{3L}^{1+\mathrm{II}} + I_{3R}^{1+\mathrm{II}} + \left[\alpha F_3' + \frac{\beta}{\sqrt{3}}F_8' - (\tfrac{8}{3})^{1/2}F_{15}'\right]. \tag{6}$$

Here $(I_3^{1+\mathrm{II}})_{L,R}$ denote the diagonal generators of $\mathrm{SU}(2)_{L,R}^{1+\mathrm{II}}$, while F_3', F_8', and F_{15}' are the diagonal generators of $\mathrm{SU}(4')_{L+R}$. The coefficients α and β of F_3' and F_8' are arbitrary. This results in the following charge assignments for the fermionic multiplets $\Psi_{L,R}$:

$$[Q(\alpha,\beta)] = \begin{bmatrix} \frac{2}{3}-\frac{1}{2}\alpha-\frac{1}{6}\beta & \frac{2}{3}+\frac{1}{2}\alpha-\frac{1}{6}\beta & \frac{2}{3}+\frac{1}{3}\beta & 0 \\ -\frac{1}{3}-\frac{1}{2}\alpha-\frac{1}{6}\beta & -\frac{1}{3}+\frac{1}{2}\alpha-\frac{1}{6}\beta & -\frac{1}{3}+\frac{1}{3}\beta & -1 \\ -\frac{1}{3}-\frac{1}{2}\alpha-\frac{1}{6}\beta & -\frac{1}{3}+\frac{1}{2}\alpha-\frac{1}{6}\beta & -\frac{1}{3}+\frac{1}{3}\beta & -1 \\ \frac{2}{3}-\frac{1}{2}\alpha-\frac{1}{6}\beta & \frac{2}{3}+\frac{1}{2}\alpha-\frac{1}{6}\beta & \frac{2}{3}+\frac{1}{3}\beta & 0 \end{bmatrix}. \tag{7}$$

Note that the baryonic quarks (in the first three columns) may be assigned a wide variety of charges, but leptons associated with the fourth color possess the *unique assignment* of charges[3] $(0,-1,-1,0)$. In the sequel we shall consider two special choices for α and β:

the integer-charge model: $\alpha=\beta=+1$, (8)

the fractional-charge model: $\alpha=\beta=0$, (9)

which serve to bring out the main contrasting features of different submodels within our scheme. Corresponding to the charge formula (6), the photon field will be made up of appropriate pieces from W_{3L}, W_{3R}, $V(8)$, and S^0 (see Sec. IV). [Note that SU(4) and SU(4′) contribute symmetrically to

Q for the integer-charge model, which we concentrate on, in the main, unless otherwise stated.]

D. Variants to the "basic" model

If electron number L_e and muon number L_μ correspond to *distinct* colors, the following simple variants may be considered:

(a) The "economical" model. Take as basic fermions the four 8-folds:

$$(\Psi_e)_{L,R} = \begin{pmatrix} \mathcal{P}_a & \mathcal{P}_b & \mathcal{P}_c & \nu \\ \mathcal{N}_a & \mathcal{N}_b & \mathcal{N}_c & e^- \end{pmatrix}_{L,R},$$

$$(\Psi_\mu)_{L,R} = \begin{pmatrix} \lambda_a & \lambda_b & \lambda_c & \mu^- \\ \chi_a & \chi_b & \chi_c & \nu' \end{pmatrix}_{L,R},$$

with the symmetry group

$$SU(2)_L \times SU(2)_R \times SU(4')_e \times SU(4')_\mu .$$

The number of fermions is the same as in the "basic" model; however, the number of gauge bosons has increased to $3+3+15+15=36$. The physical SU(3′) may now be identified with the diagonal sum of $SU(3')_e$ and $SU(3')_\mu$, whose emergence will require a more elaborate Higgs-Kibble set of scalars than are needed for the "basic" model (see Sec. IV).

(b) The "prodigal" model. A model similar in structure to the "basic" model, although more prodigal in quarks and leptons needed, could be constructed with the following basic fermions:

$$(\Psi_e)_{L,R} = \begin{bmatrix} \mathcal{P}_a & \mathcal{P}_b & \mathcal{P}_c & E^0 \\ \mathcal{N}_a & \mathcal{N}_b & \mathcal{N}_c & E^- \\ \lambda_a & \lambda_b & \lambda_c & e^- \\ \chi_a & \chi_b & \chi_c & \nu \end{bmatrix}_{L,R},$$

$$(\Psi_\mu)_{L,R} = \begin{bmatrix} \mathcal{P}'_a & \mathcal{P}'_b & \mathcal{P}'_c & M^0 \\ \mathcal{N}'_a & \mathcal{N}'_b & \mathcal{N}'_c & M^- \\ \lambda'_a & \lambda'_b & \lambda'_c & \mu^- \\ \chi'_a & \chi'_b & \chi'_c & \nu' \end{bmatrix}_{L,R}.$$

Here E^0, E^-, M^0, M^- are new heavy leptons and the primed particles are new quarks.[14] Notice that for this model both neutrinos (ν' as well as ν) can be "charmed" so that they may be coupled to charmed quarks χ and χ' through X. (This will have implications for the limits on masses of X particles; see Sec. III.)

(c) The five-color model. One may take as the basic set of fermions a 20-fold Ψ with the symmetry group, $SU(4)_L \times SU(4)_R \times SU(5')$, where

$$\Psi_{L,R} = \begin{bmatrix} \mathcal{P}_a & \mathcal{P}_b & \mathcal{P}_c & E^0 & M^0 \\ \mathcal{N}_a & \mathcal{N}_b & \mathcal{N}_c & E^- & M^- \\ \lambda_a & \lambda_b & \lambda_c & e^- & \mu^- \\ \chi_a & \chi_b & \chi_c & \nu & \nu' \end{bmatrix}_{L,R}.$$

Once again, we can assign "charm" both to ν and ν'. [As will be seen in Sec. III, what chiefly distinguishes all these variants from the "basic" model is the forbiddenness of the transition $K^0 \to e^- + \mu^+$. This transition is allowed in the "basic" model through the mediation of the exotic X's. As a consequence of this, while for the "basic" model the X's must be superheavy ($m_X > 10^4$–10^5 GeV), they need not be much more massive than 10^2–10^3 GeV for the variants.]

From a pure theoretical point of view, none of these "variants" is as attractive as the "basic model." However, at this state of experimental uncertainty, we do not wish to prejudice the issue of a final choice.

III. LIMITS ON GAUGE-MESON MASSES

All models discussed above give rise to exotic strong interactions. In order to account for their absence in the present energy domain, some of the gauge mesons must be heavy or superheavy. Such interactions are generated by three sets of gauge bosons in the scheme:

(1) the exotic vector X triplet $(X^0, X^-, X^{-\prime})$, whose interactions in the "basic" model read [see (5) and (16a)]

$$f[X^0(\bar{\nu}\mathcal{P}_a + \bar{e}\mathcal{N}_a + \bar{\mu}\lambda_a + \bar{\nu}'\chi_a) + X^-(\bar{\nu}\mathcal{P}_b + \bar{e}\mathcal{N}_b + \bar{\mu}\lambda_b + \bar{\nu}'\chi_b) + X^{-\prime}(\bar{\nu}\mathcal{P}_c + \bar{e}\mathcal{N}_c + \bar{\mu}\lambda_c + \bar{\nu}'\chi_c) + \text{H.c.}]\,; \quad (10)$$

(2) the exotic S^0 meson, whose coupling is given by

$$\frac{-f}{\sqrt{24}}\Big[\sum_{\alpha=a,b,c}(\bar{\mathcal{P}}_\alpha\mathcal{P}_\alpha + \bar{\mathcal{N}}_\alpha\mathcal{N}_\alpha + \bar{\lambda}_\alpha\lambda_\alpha + \bar{\chi}_\alpha\chi_\alpha) - 3(\bar{\nu}\nu + \bar{e}e + \bar{\mu}\mu + \bar{\nu}'\nu')\Big]S^0\,; \quad (11)$$

(3) the right-hand gauge mesons W_R, which lead to weak $(V+A)$ interactions.

The following sequences of masses will suppress reactions arising from sources (1), (2), and (3) to the presently observed extent, both in tree and (one can show) also for the loop diagrams (see Sec. IV D for an example of the operation of the suppression mechanism for loop diagrams[15]).

(i) The X couplings contribute to $\eta^0, \pi^0 \to e^+e^-$, $\mu^+\mu^-$, and (in the "basic" model *only*) to $K^0 \to e^- + \mu^+$, $\bar{K}^0 \to e^+ + \mu^-$. Since the observed amplitude for $K_L \to \mu^+ + \mu^-$ is of the order of $G_F\alpha^2$ and

no events of the variety $K_L \to \mu^\pm + e^\mp$ have yet been observed, there is a lower limit on the mass of X in the "basic" model given by $f^2/m_X^2 < G_F\alpha^2$. For $f^2/4\pi \approx 1$, this implies that X must be superheavy ($m_X > 3\times10^4$ GeV). For variants to the "basic" model, where $K \to e^- + \mu^+$ is forbidden, X need not be much more massive than $m_X^2 \approx G_F^{-1}f^2$, the severest lower limit on m_X coming from nuclear β decay and the ν_e hadronic interaction in the "economical" model. (Note that the ν_μ hadronic interactions through the mediation of X particles are suppressed in the "economical" model since known hadrons are basically charmless and ν_μ carries charm. In the "prodigal" model, both ν_e and ν_μ are charmed so that the lower limit restrictions on m_X are even less severe.)

In the following section, we exhibit the scheme for generating masses of the gauge bosons for the "basic" model only. It is worth remarking, however, that a nonsuperheavy X ($m_X \simeq 100$ GeV in the "prodigal" model) will influence $e^-e^+ \to$ hadrons at present center-of-mass energies $\simeq 5$ GeV and may provide an explanation[16] of the recently observed near constancy of the annihilation cross section over a wide range of energies.

(ii) The S^0 coupling leads to order f^2 interactions of neutrinos with hadrons and leptons. In order that the effective strength of such interactions is less than or of order G_{Fermi} at low energies, we expect $(f^2/m_{S^0}^2) \lesssim G_F$.

(iii) From the presently observed helicities and other weak-interaction experiments it appears that the $V+A$ amplitudes are at most of order 10% of $V-A$ amplitudes, from which we may conclude (*if* $g_L \simeq g_R$) that $m_{W_R^\pm} \gtrsim 3m_{W_L^\pm}$.

In addition to the restrictions on the masses of the exotic gauge bosons, there are constraints on the masses of $W_L^\pm$ and the color octet $V(8)$, due to the fact that they should mediate the known $V-A$ interactions and effective strong interactions (between baryonic quarks), respectively. From this we expect that

$$(m_{W_L^\pm})^2 \gtrsim G_F^{-1}\alpha, \qquad m(V(8)) \simeq 3\text{–}10 \text{ GeV}. \tag{12}$$

We present a summary of expected masses for gauge particles in Table I.

IV. SPONTANEOUS-SYMMETRY-BREAKING MECHANISM

A. Higgs-Kibble particles

In order to generate the postulated sequences of gauge masses (as well as Fermi masses)—and even more important, in order to motivate the broken symmetries observed in nature [i.e., global SU(3), or rather $SU_I(2)\times U_Y(1)$ when $g=0$, and the Cabibbo rotation when $g\neq 0$]—one is obliged, until a new renormalizable mechanism is invented, to implement spontaneous symmetry breaking through the expectation-value mechanism of Higgs-Kibble scalar multiplets. (We expect the situation will change with the advent of new ingredients, which may eliminate the need for such scalars, except as a means for bookkeeping in the orderly emergence of the symmetry-breaking pattern.)

At the present state of the art, there is a considerable degree of flexibility in the choice of basic Higgs-Kibble multiplets. However, it is good to reemphasize that once these multiplets are chosen and their general invariant renormalizable (cubic or quartic) interaction potential written down, the pattern of (lowest-order) symmetry breaking which emerges on minimizing this potential is (as a rule) fairly restrictive. This pattern may, of course, get drastically modified through the (radiatively generated) higher-order terms in the effective potential, as shown by Coleman and Weinberg.[17] However, as a first orientation the demand that this particular lowest-order pattern correspond fairly to the physically observed pattern of broken symmetries, or at least to a set of natural symmetries, radiative deviations from which are in principle calculable, may make some choices of basic Higgs-Kibble multiplets more desirable than others.

Be that as it may, a simple choice, capable of satisfying the restrictions on the gauge meson masses for the "basic" model discussed in Sec. III, is provided by a set of three 16-fold complex multiplets, with the cyclic transformation properties:

TABLE I. Summary of expected masses for gauge particles.

Particle	Coupling	Expected $(\text{mass})^2$
$V(8)$	$\frac{f^2}{4\pi} \simeq 1\text{–}10$	$\simeq \alpha^2 G_F^{-1}$
W_L	$\frac{g_L^2}{4\pi} \simeq \alpha = \frac{1}{137}$	$\gtrsim 4\pi\alpha G_F^{-1}$
W_R	$\frac{g_R^2}{4\pi} \simeq \alpha = \frac{1}{137}$	$> (3m_{W_L^\pm})^2$
S^0	$\frac{f^2}{4\pi} \simeq 1\text{–}10$	$> f^2 G_F^{-1}$
X	$\frac{f^2}{4\pi} \simeq 1\text{–}10$	superheavy $> \alpha^{-2}G_F^{-1}$ "basic" model heavy $> f^2 G_F^{-1}$ "economical" model heavy $\gtrsim f^2\alpha G_F^{-1}$ "prodigal" model

$$A=(4,\bar{4},1)_G\to U_L A U_R^{-1},$$
$$B=(1,4,\bar{4})_G\to U_R B V^{-1}, \qquad (13)$$
$$C=(\bar{4},1,4)_G\to V C U_L^{-1},$$

where U_L, U_R, V refer to the three global groups $U(4)_L$, $U(4)_R$, and $U(4')_{L+R}$, respectively.

The most general renormalizable quadratic and quartic potential[18] for the three multiplets $V(A,B,C)$ invariant under $U(4)_L\times U(4)_R\times U(4')$ contains twelve parameters [fifteen if the global group is specialized to $SU(4)_L\times SU(4)_R\times SU(4')_{L+R}$] besides the three mass parameters for A, B, and C, provided we impose on the theory the discrete symmetry $A\to -A$, $B\to -B$, $C\to -C$, and $\psi\to\gamma_5\psi$. One can now show that this 12-parameter potential possesses a minimum,[19] provided that $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$ are of the form

$$\langle A\rangle = R(\theta_L)R(\phi_L)\langle A_D\rangle R^{-1}(\theta_R)R^{-1}(\phi_R),$$
$$\langle B\rangle = R(\theta_R)R(\phi_R)\langle B_D\rangle, \qquad (14)$$
$$\langle C\rangle = \langle C_D\rangle R^{-1}(\theta_L)R^{-1}(\phi_L).$$

Here $R(\theta)$ and $R(\phi)$ are "Cabibbo rotations" of angles θ and ϕ in the $(\mathfrak{N},\lambda)$ and $(\mathcal{P},\chi)$ spaces, respectively, i.e.,

$$R(\theta)=\begin{bmatrix}1&0&0&0\\0&\cos\theta&\sin\theta&0\\0&-\sin\theta&\cos\theta&0\\0&0&0&1\end{bmatrix}, \text{ etc.}$$

while $\langle A_D\rangle$, $\langle B_D\rangle$, and $\langle C_D\rangle$ are diagonal and of the form

$$\langle A_D\rangle=\begin{bmatrix}a_1&&&\\&a_1&&\\&&a_1&\\&&&a_4\end{bmatrix},$$
$$\langle B_D\rangle=\begin{bmatrix}0&&&\\&0&&\\&&0&\\&&&b_4\end{bmatrix}, \qquad (15)$$
$$\langle C_D\rangle=\begin{bmatrix}c_1&&&\\&c_1&&\\&&c_1&\\&&&c_4\end{bmatrix}.$$

The four angles $\theta_{L,R}$ and $\phi_{L,R}$ are arbitrary at this stage, but the five parameters a_1, a_4, b_4, c_1, c_4 are fully determined in terms of the fifteen parameters of $V(A,B,C)$. Note the remarkable emergence of a global U(3) as the residual symmetry at this stage. [For the 15-parameter potential invariant for $SU(4)_L\times SU(4)_R\times SU(4')$, an extremum exists with

$$A_D=\begin{bmatrix}a_1&&&\\&a_1&&\\&&a_3&\\&&&a_4\end{bmatrix},$$

where $a_1\neq a_3$, so that there is a possibility of residual symmetry being in fact $SU(2)\times U(1)$.]

B. The Lagrangian of the model

Consider now the Lagrangian for the "basic" model

$$-\mathfrak{L}=\mathrm{Tr}\Big[\sum_{L,R}(\bar{\Psi}\nabla\!\!\!/\Psi)+\sum_{A,B,C}|\nabla A|^2+\sum_{V,W_L,W_R}|\nabla V|^2 + V(A,B,C)+\mu\bar{\Psi}_L A\Psi_R+\mathrm{H.c.}\Big], \qquad (16a)$$

where

$$\nabla\!\!\!/\Psi_{L,R}=\gamma_\mu(\partial_\mu\Psi+igW_\mu\Psi-if\Psi V_\mu)|_{L,R},$$
$$\nabla A=\partial A+ig_L W_L A-ig_R A W_R,$$
$$\nabla B=\partial B+ig_R W_R B-ifBV,$$
$$\nabla C=\partial C+ifVC-ig_L C W_L.$$

Barring the W-containing terms ($g_L=g_R=0$), this Lagrangian is invariant for the full symmetry $G=SU(4)_L\times SU(4)_R\times SU(4')_{L+R}$.

For $g_L\neq 0$, $g_R\neq 0$, invariance holds for the *local* subgroup $\mathcal{G}=SU(2)_L^{I+II}\times SU(2)_R^{I+II}\times SU(4')_{L+R}$, with Ψ_L, Ψ_R, A, B, and C transforming[20] as $(2+2,1,\bar{4})$, $(1,2+2,\bar{4})$, $(2+2,2+2,1)$, $(1,2+2,\bar{4})$, and $(2+2,1,4)$ representations of $\mathcal{G}$. The only new feature, so far as invariance for $\mathcal{G}$ is concerned, is that in addition to the terms included in $V(A,B,C)$ and $\mu\,\mathrm{Tr}[\bar{\Psi}_L A\Psi_R]$, one could now write a host of new renormalizable couplings among these fields:

$$-\delta\mathfrak{L}=\lambda\delta V(A,B,C) + \sum_i \mu_{ij}\,\mathrm{Tr}[\bar{\Psi}_L\Gamma_i A\Gamma_j\Psi_R]+\mathrm{H.c.}, \qquad (16b)$$

where the Γ_i's are numerical matrices. Such terms are invariant for the subgroup $\mathcal{G}$ but not for G, and act as a perturbation $\delta\mathfrak{L}$ to $\mathfrak{L}$.

If the local subgroup $\mathcal{G}$ we are dealing with were $Sp(4)_L\times Sp(4)_R\times SU(4')$ [or even $(SU(2)^I\times SU(2)^{II})_L\times(SU(2)^I\times SU(2)^{II})_R\times SU(4')$], one may prove an important result about the minimization of $V+\lambda\delta V$. This states that the minimization of $V+\lambda\delta V$ leads to solutions for $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$, which in the limit $\lambda\to 0$ reduce to the unperturbed solutions given (for example) by (15). We have verified the result by examining the detailed structure of the "perturbation" term $\lambda\delta V(A,B,C)$. We conjecture that the same result holds for the local subgroup of in-

terest here, viz. $SU(2)_L^{I+II} \times SU(2)_R^{I+II} \times SU(4')$. If true, this would imply that the (Cabibbo) angles θ, ϕ, etc., as well as departures from the residual global symmetry SU(3), are noncatastrophic functions of λ (and of the radiative corrections to λ, of order g^2, $g^2 f^2$, λf^2, etc.). This has the consequence that by ignoring $\delta\mathcal{L}$ in the first instance [for small or zero renormalized λ, and with the neglect of $O(g^2)$ radiative corrections], we are not running the risk of losing out in unexpected physics so far as the pattern of symmetry breaking is concerned. In view of this we shall henceforth drop the $\delta\mathcal{L}$ terms and work with $\mathcal{L}$. This implies that we expect all further breakdown of symmetries to be radiative in origin.

C. The mass matrix

Returning to the Lagrangian $\mathcal{L}$, let us study the mass terms, obtained by replacing A, B, C by $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$. It is convenient to define the physical Fermi fields which diagonalize the Fermi mass matrix through the relation

$$\Psi = R(\theta)R(\phi)\Psi_D .$$

In terms of Ψ_D, and the fields A_D, B_D, C_D [defined through relations similar to (14)], we can write $\mathcal{L}$ in the form

$$-\mathcal{L} = \mathrm{Tr}\Big[\sum_{L,R}(\bar{\Psi}_D \nabla_D \Psi_D) + \sum_{A,B,C}|\nabla_D A_D|^2 + \sum_{V,W_L,W_R}|\nabla V|^2 + V(A_D, B_D, C_D) + \mu(\bar{\Psi}_L A \Psi_R)_D + \mathrm{H.c.}\Big]. \tag{16c}$$

Here

$$\nabla_D \Psi_D = \partial\Psi_D + igW(\theta,\phi)\Psi_D - if\Psi_D V ;$$

$$\nabla_D A_D = \partial A_D + ig_L W_L(\theta,\phi)A_D - ig_R A_D W_R(\theta,\phi),$$

etc.

with

$$W(\theta,\phi) = R^{-1}(\theta)R^{-1}(\phi)WR(\theta)R(\phi) .$$

We wish now to consider the gauge-meson sector of the mass matrix, including the mixing terms, in more detail. For a first orientation, take $\theta_L = \phi_L = \phi_R = \theta_R = 0$. This leads [with the usual replacement of A by $(A + \langle A\rangle)$ in the $|\nabla_\mu A|^2$ term, and similarly for the B and C fields] to the following expression for the gauge-meson mass terms:

$$\begin{aligned} L^I_{\mathrm{mass}} = {} & \tfrac{1}{4} g_L{}^2(a^2+c^2)[2W_L^+W_L^- + (W_L^3)^2] + \tfrac{1}{4} g_R{}^2(a^2+b_4{}^2)[2W_R^+W_R^- + (W_R^3)^2] \\ & - \tfrac{1}{2} g_L g_R (a^2)(W_L^3 W_R^3) + \tfrac{1}{2} g_L g_R (a_1{}^2 + a_1 a_4)(W_L^+ W_R^- + W_L^- W_R^+) + \tfrac{1}{2} f^2 c_1{}^2 (V_3{}^2 + V_8{}^2 + 2V_\rho^+ V_\rho^- + 2V_{K*}^+ V_{K*}^- + 2V_{K*}^0 \bar{V}_{K*}^0) \\ & + \tfrac{1}{6} f^2 (c_1{}^2 + 3c_4{}^2 + 3b_4{}^2)(S^0)^2 + \tfrac{1}{2} f^2 (c_1{}^2 + c_4{}^2 + b_4{}^2)(X^0\bar{X}^0 + X^- X^+ + X^{-\prime} X^{+\prime}) - f g_L W_L^3 \Big[c_1{}^2 \Big(V_3 + \frac{V_8}{\sqrt{3}} - \frac{S^0}{\sqrt{24}} \Big) + \frac{3c_4{}^2}{\sqrt{24}} S^0 \Big] \\ & + f g_R \frac{3b_4{}^2}{\sqrt{24}} (W_R^3 S^0) + f g_L (c_1{}^2)(W_L^+ V_\rho^- + W_L^- V_\rho^+) + f g_L (c_1 c_4)(W_L^+ X^{-\prime} + W_L^- X^{+\prime}), \end{aligned} \tag{17}$$

where

$$a^2 \equiv 3a_1{}^2 + a_4{}^2 ,$$

and

$$c^2 = 3c_1{}^2 + c_4{}^2 .$$

These give rise to the following masses for the gauge bosons (note that for *this* purpose one may safely ignore the mixing terms such as $W_R^+ V_\rho^-$, $W_R^+ X^{-\prime}$, and even $W_L^+ W_R^-$, a fact which is better justified *a posteriori*; of course, we do not neglect the important physical consequences of the mixing terms):

$$m_{W_L^\pm} \simeq \tfrac{1}{2} g_L (a^2 + c^2)^{1/2} ,$$

$$m_{W_R^\pm} \simeq \tfrac{1}{2} g_R (a^2 + b_4{}^2)^{1/2} ,$$

$$m(V(8)) \simeq \frac{f c_1}{\sqrt{2}} ,$$

$$m(X^0, X^-, X^{-\prime}) \simeq \frac{f(c_1{}^2 + c_4{}^2 + b_4{}^2)^{1/2}}{\sqrt{2}} , \tag{18}$$

$$m_{S^0} \simeq \frac{f}{2\sqrt{2}} (c_1{}^2 + 3c_4{}^2 + 3b_4{}^2)^{1/2} ,$$

$$m_{A^0} = 0 ,$$

$$m_{Z^0} \simeq \tfrac{1}{2} (g_L{}^2 + g_R{}^2)^{1/2} a .$$

Here A^0 denotes the photon and Z^0 a neutral eigenstate, whose complexion is exhibited later. From these expressions one may infer that the restrictions on the gauge-meson masses for the "basic" model outlined in Sec. III are satisfied if we assume that there are basically *three essentially different scales of masses* (vacuum expectation values), characterized by

$$\begin{aligned} & c \simeq (c_1, c_4) \lesssim 1 \text{ GeV} , \\ & a \simeq (a_1, a_4) \approx 300 \text{ GeV} , \\ & b = b_4 \approx 10^4 - 10^5 \text{ GeV} . \end{aligned} \tag{19}$$

The above pattern of vacuum expectation values has the consequence that members of the SU(3′) octet of gauge mesons $V(8)$ possess nearly equal masses of few GeV; $W_L^{\pm}$ and Z^0 have masses of order 100 GeV; *while all*[21] *the exotic gauge mesons* ($X^0, X^-, X^{-\prime}, S^0$, and $W_R^{\pm}$) acquire (heavy or) superheavy masses because of the *single parameter* b_4 being large. Looked upon from this point of view, baryon-lepton asymmetry and left-right asymmetry in the low-energy domain is due to this new scale of mass b_4. [Note that *a priori* the roles of b_4 and c_4 are interchangeable for most purposes except for baryon-lepton number violation. One may remark that for the "prodigal" model (which we have not treated in any detail here) b_4 could be as small as 10^2–10^3 GeV, while[19] $c_4 \approx c_1\alpha^2$.]

It is straightforward to diagonalize the vector-meson mass matrix. For the sake of facilitating further discussions, we show below the composition of only the neutral "diagonal" fields. These are obtained by consistently neglecting terms of order (c^2/a^2), (c^2/b^2), and $(g_{L,R}{}^2/f^2)$, whenever such terms are of no physical significance[22]:

$$
\begin{aligned}
&\frac{A}{e} = \left(\frac{W_L^3}{g_L} + \frac{W_R^3}{g_R}\right) + \frac{(V_3 + V_8/\sqrt{3} - (\frac{2}{3})^{1/2}S^0)}{f}, \quad m_A = 0 \\
&Z^0 = \frac{f(g_R W_R^3 - g_L W_L^3) - (\frac{2}{3})^{1/2} g_R{}^2 S^0 + O(g^2/f^2, c^2/a^2)}{[f^2(g_R{}^2 + g_L{}^2) + \frac{2}{3}g_R{}^4]^{1/2}}, \quad m_{Z^0} \simeq \tfrac{1}{2}(g_L{}^2 + g_R{}^2)^{1/2} a \\
&S^0 = \frac{fS^0 + (\frac{2}{3})^{1/2} g_R W_R^3 + O(g^2/f^2, c^2/a^2)}{[f^2 + \frac{2}{3}g_R{}^2]^{1/2}}, \quad m_{S^0} \simeq (\tfrac{8}{3})^{1/2} f b_4 \\
&U^0 = \frac{g_R W_L^3 + g_L(1+\Delta) W_R^3 - \sqrt{3}\,\bar{f} U^0 - (\frac{2}{3})^{1/2}(g_L g_R/f)(1 + \Delta + \frac{4}{3}c_1{}^2/b_4{}^2)S^0}{[g_R{}^2 + g_L{}^2 + 3\bar{f}^2 + \frac{2}{3}(g_L g_R/f)^2]^{1/2}}, \quad m_{U^0} \simeq \frac{f c_1}{\sqrt{2}}
\end{aligned}
\tag{20}
$$

and

$$
V^0 = \tfrac{1}{2}(V_3 - \sqrt{3}\,V_8), \quad m_{V^0} = \frac{f c_1}{\sqrt{2}},
$$

where[23]

$$
\begin{aligned}
&\frac{1}{e^2} = \frac{1}{g_L{}^2} + \frac{1}{g_R{}^2} + \frac{2}{f^2}, \\
&U^0 = \tfrac{1}{2}(\sqrt{3}\,V_3 + V_8), \\
&\Delta = 2(f^2/g^2)(c_1{}^2/a^2), \\
&\bar{f} = \frac{f(g_R{}^2 + g_L{}^2)}{2 g_R g_L}.
\end{aligned}
\tag{21}
$$

D. Color-valency mixing

Note the important circumstance that as a consequence of gauging the B and C multiplets, color and and valency mix,[24] and in particular the exotic X's mix with the W's, leading to a nonconservation of baryon (and lepton) numbers. This mixing term in the mass matrix equals $f g_L c_1 c_4 (W_L^{\dagger} X^{-\prime} + \text{H.c.})$. Note the following features of this term:

(a) The strength of baryon-lepton number $(B-L)$ violating interaction is directly proportional to c_4, with exact conservation[25] obtaining for $c_4 = 0$.

(b) The W-X mixing term responsible for baryon violation gives rise to the effective propagator

$$
\langle WX \rangle = \frac{f g c_1 c_4}{(k^2 + m_X{}^2)(k^2 + m_W{}^2)}
$$

in momentum space. This propagator is highly convergent so that no infinities are ever encountered in closed-loop calculations involving W-X mixing. [Using standard arguments, i.e., diagonalization of the W-X mass matrix, it is easy to see that the effective coupling strength for the baryon-lepton number violation is $\simeq (f^2 g^2 c_1 c_4)/m_X{}^2$ so far as closed-loop contributions are concerned.]

E. Gauge-meson masses for the fractional-charge model

In this case, it is worth noting that simple representations of Higgs-Kibble multiplets (for example, B and C) *cannot* be utilized to give masses[26] to the $V(8)$ octet of gauge mesons, though they can furnish masses for the X's and S^0. This is because the appropriate entries in the multiplet, capable of giving masses to $V(8)$, carry electric charge (if $\alpha = \beta = 0$) and therefore must possess zero vacuum expectation values. One can, however, introduce higher *reducible* multiplets such as $(1, 1, 4\times4\times4)$ if one wishes to give masses to the octet SU(3′) gauge mesons. [Note that these restrictions do not apply if (unlike the case for the present scheme) the electric charge contain contributions from an Abelian U(1) gauge.] Note also that the X's in this scheme are fractionally charged so that the X's and W's can never mix. In other words, in a model of the type described above, baryon-lepton number conservation is a conse-

quence of the twin postulates of fermion number and electric charge conservation.

V. FERMION MASSES

A. The fermion mass term

The fermion mass term

$$\mu\,\mathrm{Tr}[(\bar{\Psi}_L\langle A\rangle\Psi_R)_D+\mathrm{H.c.}],$$

with

$$\langle A_D\rangle=\begin{bmatrix} a_1 & & & \\ & a_1 & & \\ & & a_3 & \\ & & & a_4 \end{bmatrix}$$

cannot provide a distinction between fermions of different colors. Thus baryonic quarks and leptons in the same row of Ψ possess the same mass (in particular $e\leftrightarrow\mathcal{N}$, $\mu\leftrightarrow\lambda$). The situation is not remedied by the possibility [see (16b)] of adding Yukawa couplings of the type

$$\mu_{ij}\,\mathrm{Tr}(\bar{\Psi}_L\Gamma_i A\Gamma_j\Psi_R)\,.$$

Such couplings [which in effect treat A as constituted of four independent submultiplets[20] (2, 2, 1) of $SU(2)_L\times SU(2)_R\times SU(4')$] may assign different masses to $\mathcal{P}$, $\mathcal{N}$, λ, and χ within the same column. However, the SU(4')-singlet character of the submultiplets means that there still would be no color distinction.

Such a color distinction could arise if we were willing to introduce an SU(4')-nonsinglet scalar multiplet such as $A'=(4,\bar{4},15)_G$ [or a smaller multiplet $(2,2,15)_G$], which is capable of having gauge-invariant Yukawa coupling with fermions. This would have the consequence of assigning one mass (call it a') to the three baryonic quarks in a given row and $-3a'$ to the lepton in the same row of Ψ. Thus the multiplets A and A' could collaborate[27] to make the baryonic quarks consistently heavy and the leptons light. We have, however, avoided[27] introducing a multiplet such as A'. Apart from the undesirability of proliferating Higgs-Kibble multiplets, we wish to retain the SU(4') color group as a "natural" symmetry in the sense that deviations from it (including baryon-lepton number violation) should eventually be computable.

Even without an A', there exists a definite mechanism in the scheme which could boost the masses of the baryonic quarks without boosting the masses of the leptons. In the strong-interaction sector the baryonic quarks are coupled to the light gauge mesons $V(8)$ as well as to the superheavy ones, whereas the leptons are coupled only to the superheavy or heavy gauge mesons (X's and S^0). Thus, the former may get most of their mass through self-energy contributions involving the *light* $V(8)$ exchanges—something not available to the leptons. Note that the mass difference (m_q-m_l) is *computable* in the scheme, with SU(4') as a natural symmetry.[28] Of course, second-order perturbation in f is not reliable; it is amusing, however, that one gets[15] the right sign and roughly the correct order of magnitude:

$$\frac{m_q-m_l}{m}\simeq\frac{3}{4\pi}\left(\frac{f^2}{4\pi}\right)\ln\left(\frac{m_X^{\,2}}{m_{V(8)}^{\,2}}\right)\,, \tag{22}$$

which, using renormalization-group ideas, may possibly be an approximation to $\simeq(m_X^{\,2}/m_{V(8)}^{\,2})^{3f^2/(4\pi)^2}-1$. Here m is the zeroth-order common mass of quarks and leptons which may be $\approx m_l$. The baryon-lepton mass difference may quite possibly have its origin in the large magnitude for m_X.

B. The massless neutrinos

In the theory developed so far, the neutrinos ν and ν' are 4-component objects, and even if one could arrange zero bare masses for them (by introducing the multiplet A', for example), nothing can stop their acquiring mass through radiative corrections (e.g., through the γ_5-noninvariant vector interaction[29] $\bar{\nu}\gamma_\mu\mathcal{P}X_\mu$). If the physical neutrinos $(\nu_e)_L$ and $(\nu_\mu)_L$ are indeed (2-component) *massless* objects, the model is presented with a dilemma of massive neutrinos.

Below we suggest a mechanism to resolve this dilemma. The idea is this: A physical spin-$\frac{1}{2}$ particle is massless only if it can be described by a 2-component spinor. To implement the 2-component nature of the physical neutrino, postulate in the model two additional 2-component fermions, ζ_L^e and ζ_L^μ, which are singlets (1, 1, 1) of the *local* group $\mathcal{G}$, and thus possess no gauge interaction. The *only* renormalizable invariant interaction they can possess is of the Yukawa type:

$$h\bar{\zeta}_L^\mu\mathrm{Tr}[\bar{B}\Psi_R]+h'\bar{\zeta}_L^e\mathrm{Tr}[\bar{B}\Gamma\Psi_R+\mathrm{H.c.}]\,, \tag{23}$$

where the matrix

$$\Gamma=\begin{pmatrix} & \tau_1 \\ \tau_1 & \end{pmatrix}$$

commutes with the generators of $\mathcal{G}$.

Consider now the mass matrix for the complex ζ_L^μ, ν_L', ν_R'. Assuming $\theta_{L,R}=\phi_{L,R}=0$, for simplicity of discussion and diagonalizing the relevant terms:

$$hb_4\bar{\zeta}_L^\mu\nu_R'+\mu a_4\bar{\nu}_L'\nu_R'+\mathrm{H.c.}\,,$$

we immediately see that among the 6-component complex (ν_L', ν_R', and ζ_L^μ) there is one massless

2-component left-handed particle which we identify with the physical ν^μ:

$$\nu^\mu_{\text{physical}} = \frac{\mu a_4 \zeta^\mu_L - h b_4 \nu'_L}{(\mu^2 a_4^2 + b_4^2 h^2)^{1/2}} , \tag{24}$$

plus a massive 4-component fermion. Likewise for the complex $(\nu_L, \nu_R, \zeta^e_L)$. (The restrictions imposed on the parameters h, h' and the angles θ, ϕ by the demands of the μ-e universality can easily be worked out and will not be exhibited here.)

Since the entire set of couplings of ζ^μ and ζ^e are contained in (23), these fields may be integrated out for processes not involving them as incoming or outgoing particles, leaving us with an effective interaction

$$\mathcal{L}_{\text{eff}} \sim h^2 \mathrm{Tr}(\bar{B}\Psi_R)\frac{1}{\not p}\mathrm{Tr}(\bar{\Psi}_R B) + (h')^2 \mathrm{Tr}(\bar{B}\Gamma\Psi_R)\frac{1}{\not p}\mathrm{Tr}(\bar{\Psi}_R \Gamma B) . \tag{25}$$

It is amusing that $\mathcal{L}_{\text{eff}}$ does not contribute to processes involving the physical (left-handed) massless neutrinos, although it would contribute (with strength $\sim h^2$) to semileptonic processes like $e^+ + e^- \to$ hadrons, its characteristic feature being its long-range character (on account of the factor $1/\not p$) and the appearance of right-polarized particles Ψ_R.

Before closing, we remark that if radiative corrections to $\langle B \rangle$ are taken into account, such that after including these $\langle B \rangle$ has the most general form (consistent with conservation of charge)

$$\langle B \rangle = \begin{bmatrix} b_1 & & & f_1 \\ & b_2 & f_2 & \\ & f_3 & b_3 & \\ f_4 & & & b_4 \end{bmatrix} , \tag{26}$$

then the interaction term (23) will read

$$h\bar{\zeta}^\mu_L (b_1 \mathcal{P}_s + b_2 \mathfrak{N}_b + b_3 \lambda_c + b_4 \nu')_R + \text{terms containing } \zeta^e_L \text{ and } f\text{'s} . \tag{27}$$

Clearly, (27) will contribute further to a mixing of different colors and valencies, this mixing being proportional to the (small) parameter h. Conversely, the masslessness of the physical neutrinos [which led to the introduction of the interaction (23) in the first place and the necessity for $\langle B \rangle \neq 0$] may possibly rank as the deeper primary reason behind the mixing of colors and valencies and thus for the violation of *internal* symmetries observed in nature.

VI. EXPERIMENTAL CONSEQUENCES

In this section we list some of the experimental consequences of our scheme.

A. The color octet of vector gluons $V(8)$

For energies sufficiently above threshold, one may expect to produce these particles (expected masses $\approx$3–10 GeV) in *pairs* in normal hadronic collisions with reasonable cross sections. Depending on the nature of conservation of color quantum numbers, one or more of these particles may be semistable. The whole octet is electrically neutral for the fractionally charged quark model ($\alpha = \beta = 0$), while some of its members carry unit charge for the integer-charge quark model.

B. The color components of the photon (U^0)

If the photon contains a color octet component U^0 (i.e., quarks carry integer charges), the object "U^0" can be produced *singly* in photon-induced reactions such as

$$\gamma + p \to p + \text{"}U^0\text{"} , \qquad e^+ + e^- \to \text{"}U^0\text{"} . \tag{28}$$

We expect that the production cross sections for "U^0"[30] in either reaction should be comparable to that of ρ^0 at appropriate energies. The production of "U^0" in (e^-e^+) annihilation should exhibit a resonant structure similar to that of ρ^0 except that its expected total width is uncertain. If there exist color octet states[31] C' lighter than "U^0," this object would decay strongly to (C' + hadron) or ($C' + C'$); otherwise its primary decay mode would be $(\pi^0 + \gamma)$, with secondary decays to (e^+e^-) and $(\mu^+\mu^-)$. A reasonable expectation would be $\Gamma_{(\pi^0+\gamma)}$: $\Gamma_{\text{lepton pair}} \simeq 1:(10^{-3}$–$10^{-4})$. To summarize, a search for U^0 using (1) *missing-mass measurements*, (2) $(\mu^+\mu^-)$ production in $(\gamma + p)$ reactions, and (3) (e^+e^-)-annihilation experiments, may offer a direct means of establishing whether the photon contains "color" pieces. If it does, this would favor the integer-charge quark model in contrast to the model with fractional charges. In (e^-e^+) annihilation, apart from single production of "U^0"'s, one should also expect *pair production*[32] of the charged members of the spin-1 color octet mesons (in the integer-charge quark model) above the necessary threshold. It is worth emphasizing that in accordance with the light-cone or parton-model ideas, one may expect the ratio $R = \sigma(e^+e^- \to \text{hadrons})/\sigma(e^+e^- \to \mu^+\mu^-)$ to settle to a value $\sum Q_i^2 = 6$ (plus possible contributions from the charged spin-1 gluons) for the integer-charge model and to a value $\frac{10}{3}$ for the fractional-charge model. Also for the integer-charge case, due to the charged spin-1 gluon contributions in a parton model, one should expect $\sigma_{\text{longitudinal}}/\sigma_{\text{transverse}}$ to remain nonvanishing in inelastic electron-nucleon scattering.

C. Neutral neutrino-current processes

$$\nu_\mu + e \to \nu_\mu + e,$$
$$\nu_e + e \to \nu_e + e, \text{ etc. (leptonic)}, \quad (29)$$
and
$$\nu + p \to (\nu + \text{hadrons}), \text{ etc. (semileptonic)}.$$

(a) Z^0 exchange. Our Z^0 is almost identical to the "Z^{0}" of the simple $SU(2)_L \times U(1)$ gauge theories[33] insofar as *leptonic* currents are concerned. One may in fact make the correspondence[34] $(g_L, g_R) \to (g, g')$, where g and g' denote the coupling constants associated with the gauge groups $SU(2)_L$ and $U(1)$ respectively. The "Z^{0}"'s in the two schemes differ in their coupling to hadronic currents since the U^0 term in our photon does not have a counterpart in our Z^0. The differences become significant when one starts producing SU(3′) nonsinglet hadrons. This is because the U^0 current is SU(3′) octet.

(b) U^0 exchange. It is important to note that $\tilde{U}^0$ contains (W_L^3, W_R^3, and S^0) essentially in the combination

$$[g_R W_L^3 + g_L W_R^3 - (\tfrac{2}{3})^{1/2}(g_L g_R/f)S^0] .$$

This same combination also enters into the photon and is *exactly* decoupled from $(\bar{\nu}\nu)$ current. Thus $\tilde{U}^0$ contributes to leptonic processes *only through correction terms* of order Δ [see Eq. (20)], whose net effect towards leptonic amplitudes can be seen to be of order[35] $(g^2/f^2)^2(\Delta/C^2) \simeq (10^{-6}) \times$ (strong amplitude).

For semileptonic processes there is an additional contribution due to interference of U^0 and S^0 currents which leads to an amplitude of order $(g^2/f^2) \times (\Delta^2/C^2)$. Because the U^0 current is a color octet this will become significant once we are above threshold for production of color nonsinglet states.

(c) S^0 exchange. This directly contributes to both leptonic and semileptonic processes. If it is not superheavy,[21] but instead has a mass $\approx (f/g)M_{W_L^+}$, its contribution would be of order G_{Fermi}. There is thus a distinct possibility in our scheme of a *new variety of contribution*, which does not exist in the simple $SU(2)_L \times U(1)$ theory. Since the hadronic current, which is coupled to S^0, is a singlet under *both* SU(3) and SU(3′) its contribution to semileptonic processes involving low-lying hadrons is not suppressed by SU(3′) selection rules.

To summarize, a study of the cross sections of neutral neutrino-current leptonic *as well as* semileptonic processes can in the first place determine whether departures from the simple $SU(2)_L \times U(1)$-theory[33] predictions are warranted; secondly (since S^0 current is pure *isoscalar*), a study of the *isospin structure* of the hadronic current for semileptonic processes would help determine whether S^0 contribution is significantly present. We should stress that the introduction of S^0 is a consequence of our gauging of the full SU(4′) group.

D. Right-handed currents

The scheme explicitly uses $V-A$ as well as $V+A$ currents. If the W_R are not superheavy (i.e, $M_{W_R^\pm} \sim 3M_{W_L^\pm}$), one may expect $(V+A)$ amplitudes at around 10% level of the $(V-A)$ amplitudes. Observable parameters involving helicity and correlation measurements in leptonic and semileptonic weak decays, however, depend upon squares of $(V-A)$ and $(V+A)$ amplitudes without an interference term to the extent that one may neglect the masses of the neutrinos (if they are light-mass 4-component objects).

E. Anomalous interactions of e and μ

Regarding contributions from gauge mesons to such anomalous interactions, the only relevant exchange in the "basic" model is $\tilde{U}^0$, since all other gauge mesons coupled to e and μ (i.e., Z^0, S^0, and the X's) are much too heavy. One may verify [using Eq. (20)] that the $\tilde{U}^0$-exchange contribution to (ee) and $(\mu\mu)$ scattering is of order $(g^2/f^2) \times (k^2/m_U^2)$, compared to the one-photon-exchange contribution to the same process, where k^2 denotes (momentum transfer)^2; this is too small to be observable at present. By the same token, the contributions of these additional interactions to anomalous magnetic moments of the electron and the muon appear to be too small to be relevant for a present comparison of theory versus experiment. $\tilde{U}^0$ exchange would, however, contribute significantly in inelastic ep scattering once we are above the threshold for production of SU(3′) nonsinglet states. Also, so far as the variants to the "basic" model are concerned, the effects of nonsuperheavy X's could be significant for semileptonic processes[16] and particularly for $e^+ + e^- \to$ hadrons.

F. Parity violation

Parity and strangeness violations in hadronic processes arising from radiative corrections with W_L loops are computable and of order G_{Fermi} (similar to the situation for the gauge model of paper I and for exactly the same reasons)[36]; the contributions of W_R loops are suppressed by the heavier W_R mass. Furthermore, it is easy to show that no large parity and strangeness violations occur in tree diagrams involving the "diagonalized" gauge-meson fields. The one field which might have caused concern is $\tilde{U}^0$, since it is "light." However, $\tilde{U}^0$ contains W_L^3 and W_R^3 essentially in the

combination $(g_R W_L^3 + g_L W_R^3)$, which is coupled to pure vector strangeness-conserving current. Thus parity violation due to $\bar{U}^0$ exchange can arise only through a term of order Δ in the coefficient of W_R^3. Such a term interfering with the U^0 term in $\bar{U}^0$ leads to nonleptonic amplitudes of order $(g^2/f^2)\Delta \lesssim 10^{-4}$ (compared to normal hadronic amplitudes) and, when interfering with the S^0 term, leads to amplitudes $\simeq (g^2/f^2)\Delta^2 \lesssim 10^{-5}$. The fact that for the U^0 case we are dealing with an SU(3′)-octet transition operator means that there is a further suppression factor for processes involving low-lying hadrons.

G. Violation of baryon and lepton numbers

The present gauge scheme can lead—through a spontaneous-symmetry-breaking mechanism—to a violation of baryon and lepton numbers in the integer-charge quark model. This would lead to quarks disintegrating into leptons. Even with a reasonably large strength for $q \to l + l + \bar{l}$ decays $(G_q \simeq 10^{-9} m_p^{\,2})$, there is, however, no conflict[2] with the extraordinary stability of the proton, since the latter is a three-quark composite $(B=3)$ and its triple B decay $(|\Delta B| = 3)$ can occur only in order G_q^3 or higher (depending on additional selection rules). This is assuming that both the quark and the diquark systems are more massive than the proton, or even, alternatively, that there is some field-theoretic mechanism of quark imprisonment, which does not permit hadronic quarks to materialize as physical particles. [In Amati's graphic phrase, in this model, "for a quark imprisoned within a nucleon, the price of liberty is death" (Aix-En-Provence Conference, 1973).]

If G_q is deduced from the proton's lifetime, there are uncertainties in the determination of its precise value. These arise from the fact that the proton decay may be subject to additional selection rules due to SU(3′) quantum numbers (see Ref. 2) and also from the details of the wave function of the proton considered as a three-quark composite. Thus from a purely phenomenological point of view, one may search for (integer-charge) quark decays (for example, in process $q \to l + \pi$ or $q \to l + l + \bar{l}$) with lifetimes ranging from 10^{-10} to 10^{-6} sec on the one hand and very much shorter than 10^{-10} sec on the other. In any case, the highly *energetic* lepton (or leptons) in the decay products should provide a characteristic signature for quark decay. Furthermore, in the decay $q \to l + l + \bar{l}$, one pair of leptons $(l\bar{l})$ coming from virtual W decay would carry longitudinal polarization exhibiting parity violation.

As far as proton decay is concerned, one must emphasize that it is essential to search for *multiparticle* decays of the proton. This is because the minimum number of particles—with fermion number conservation—into which the proton (as a three-quark composite) can decay is three neutrinos plus a pion. Thus we may expect the following decay modes:

$$\begin{aligned} p &\to 3\nu + \pi^+ \\ &\to 4\nu + e^+ \text{ or } 4\nu + \mu^+ \\ &\to 4\nu + \mu^+ + e^+ + e^-, \text{ etc.} \end{aligned} \tag{30}$$

The crucial point is that no two- or three-body decays are allowed.[37]

To conclude, the model provides a number of intriguing experimental possibilities. In particular, we urge a search for (1) color (either gauge-vector bosons or color nonsinglet states) in photon-induced reactions, (2) large isoscalar component in neutrino-induced neutral-current processes, (3) possible anomalous interactions of e and μ, specially at energies above the threshold for color production (for example in ep reactions), (4) right-handed $(V+A)$ weak interactions, (5) muon-electron number-violating transitions such as $K \to \mu + e$ decays—this is relevant for the "basic" model only, (6) "nonsuperheavy" exotic X-meson effects in semileptonic processes in general, and for $e^+ + e^- \to$ hadrons in particular (relevant for models other than the "basic"), and finally (7) baryon-lepton number nonconservation in quark and proton decays.

ACKNOWLEDGMENTS

We appreciate discussions with J. D. Bjorken, W. Franklin, S. L. Glashow, Ling-Fong Li, R. N. Mohapatra, M. A. Rashid, D. Ross, G. A. Snow, and J. Strathdee.

*Work supported in part by the National Science Foundation under Grant No. NSF GP 20709.

[1] J. C. Pati and Abdus Salam, Phys. Rev. D 8, 1240 (1973), to be referred to as I.

[2] J. C. Pati and Abdus Salam, Phys. Rev. Lett. 31, 661 (1973), to be referred to as II.

[3] If baryons and leptons belong to distinct representations of some underlying symmetry group, there is no reason why weak interactions might not have picked $V-A$ for *baryons* and *antileptons* $(\nu_e, e^+, \mu^+, \nu_\mu)$. Since charges of leptons within the same fermionic multiplet as the quarks get *fixed* due to our assumption that they are members of a common fermionic multiplet (see Sec. II), the statement that the same $\frac{1}{2}(1+\gamma_5)$ projection occurs for the two types of matter is *not* a matter of mere terminology of what one may have chosen to call leptons.

[4]The situation here is qualitatively the same as the one encountered in the unification of weak and electromagnetic interactions, with ν_μ and μ^-, for example, sharing universal gauge interactions. The fact that ν_μ exhibits *only* weak interactions (and no interaction of electromagnetic strength) at presently available energies is attributed to the massiveness of W and Z^0.

[5]By L we mean $(L_e + L_\mu)$, where L_e and L_μ are additive electron and muon numbers. It is conceivable that L_e and L_μ need to be further distinguished and correspond to distinct colors. This would give variants of the "basic" model which are briefly considered in Sec. II D.

[6]Whether the observed SU(3) is the diagonal sum of $SU(3) \times SU(3')$ or the SU(3) within our SU(4) may be left open until one begins to observe SU(3') nonsinglet states. Either possibility can be accommodated into our scheme. Note that the notation of $SU(4') \times SU(4'')$ of I is replaced by $SU(4) \times SU(4')$ in this paper.

[7]In many ways it is attractive to consider the mathematical objects $F = (\mathcal{P}, \mathcal{N}, \lambda, \chi)$ and $B = (a, b, c, d)$ as fundamental fields and the 16-fold Ψ's to be composite. [For the "five-color" model of Sec. II D (a, b, c, d) are replaced by five colors (a, b, c, d_e, d_μ).] This may provide an answer to why the relevant global symmetry group underlying our scheme is $SU(4)_L \times SU(4)_R \times SU(4')_{L+R}$, and not SU(16) or $SU(16)_L \times SU(16)_R$. It may also provide a reason for the circumstance that one set of interactions (the weak) is chiral in nature, while the other (the strong) is vector; the difference may lie in the spins ($\frac{1}{2}$ and zero) of F and B. In this picture, a binding between F and B {which need not be strong and may be $\approx \alpha$ [see H. Fritzsch and P. Minkowski, Caltech Report No. 68-421 (unpublished)]} would have to arise through some additional mechanism. We propose to pursue the consistency of this idea elsewhere (in particular, the consequences of treating the known leptons as composites). J. D. Bjorken and, independently, O. W. Greenberg (private communications) have been interested in related ideas (specially with regard to the 3-triplet model) from different motivations.

[8]Here ν and ν' are 4-component objects. Later we shall introduce two left-handed singlets ζ^e_L and ζ^μ_L to ensure the emergence of 2-component massless neutrinos ν^e_L and ν^μ_L. The "basic" model, in its final form, thus may contain a total of 16 + 16 + 2 = 34 2-component fields, if neutrinos are massless.

[9]We are indebted to Professor S. Glashow for pointing this out to us.

[10]We have in mind questions such as (i) calculability and strength of parity and strangeness violations [see S. Weinberg, Phys. Rev. Lett. 31, 494 (1973); R. N. Mohapatra, J. C. Pati, and P. Vinciarelli, Phys. Rev. D 8, 3652 (1973)], (ii) classification symmetry of hadrons being limited to $SU(3) \times SU(3')$ and no higher [see J. C. Pati and Adbus Salam, ICTP, Trieste, Internal Report No. IC/73/81 (unpublished)], (iii) saturation properties of hadrons treated as bound states of qqq and $q\bar{q}$ [see M. Han and Y. Nambu, Phys. Rev. 139, B1006 (1965); O. W. Greenberg and D. Zwanziger, Phys. Rev. 150, 1177 (1966); and H. Lipkin, SLAC report, 1973 (unpublished)], and (iv) asymptotic freedom for strong interactions [D. Gross and F. Wilczek, Phys. Rev. Lett. 30, 1343 (1973); and D. Politzer, *ibid.* 30, 1346 (1973)]. {Whether asymptotic freedom is lost by the introduction of Higgs-Kibble scalars in a model like ours based on a semisimple group like $SU(4)_L \times SU(4)_R \times SU(4')$ is discussed in a forthcoming paper by R. Delbourgo and Abdus Salam [Imperial College report (unpublished)]. At any rate this loss is unlikely to be serious if Higgs-Kibble scalars are composites.}

[11]Note that if we did not insist on the full SU(4') gauging, we could still obtain an elegant left-right symmetric anomaly-free gauge scheme given by $SU(2)^{I+II}_L \times SU(2)^{I+II}_R \times SU(3')_{L+R} \times U(1)_{L+R}$, where U(1) is given by the charm generator $(\frac{2}{3})^{1/2}(F'_{15})_{L+R}$ of SU(4'). This scheme will have all the consequences of the present extended scheme except for the interactions mediated by the X particles. [The U(1)-gauge particle will correspond to S^0 of the present scheme, though its coupling strength would not be related to that of the SU(3') octet.] In this restricted scheme, it may be possible to preserve masslessness of neutrinos without invoking ζ's (footnote 8). This is because $\nu^{e,\mu}_L$ and $\nu^{e,\mu}_R$ interactions (mediated through W_L, W_R and S^0 gauge mesons) are γ_5-invariant. (One must still examine whether the leftover Higgs-Kibble meson interactions can be arranged to preserve the said γ_5 invariance. We thank Dr. R. N. Mohapatra for pointing this out to us.)

[12]The charges of the X triplet $(0, -1, -1)$ correspond to the integer-charge model. For the fractional-charge model these would be $(-\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3})$.

[13]Note that $(I^{I+II}_{3L} + I^{I+II}_{3R})$ equals $(F_3 + F_8/\sqrt{3} - (\frac{2}{3})^{1/2}F_{15})_{L+R}$, where F_3, F_8, and F_{15} are the diagonal generators of SU(4) defined explicitly in I.

[14]The symmetry group is $SU(4)_L \times SU(4)_R \times SU(4')$. It could also be $SU(4)_L \times SU(4)_R \times SU(4')_e \times SU(4')_\mu$, in which case the μ-e distinction is emphasized even more.

[15]The problems of calculability (and renormalizability) of the relevant radiative corrections in the present model have been considered by Dr. D. Ross [Imperial College report (unpublished)].

[16]This suggestion (first made at the Irvine Conference on Weak Interactions, 1973) is the subject of a separate paper [J. C. Pati and Abdus Salam, Phys. Rev. Lett. 32, 1083 (1974)]. *Note added in proof.* However, it should be stressed that in the gauge-theory context, we find it difficult to find a suitable Higgs-Kibble system, which in zeroth order can assign light mass of order 100 BeV to x's and yet appropriately heavy mass to s.

[17]S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973).

[18]The general 15-parameter potential $V(A,B,C)$ contains mass terms like $\mathrm{Tr}[AA^\dagger]$ and terms of the type $(\mathrm{Tr}[AA^\dagger])(\mathrm{Tr}[BB^\dagger])$, $\mathrm{Tr}[CAA^\dagger C^\dagger]$, etc., and is invariant for $U(4)_L \times U(4)_R \times U(4')$. When one specializes to $SU(4)_L \times SU(4)_R \times SU(4')$, three additional terms appear. These are proportional to $\det A$, $\det B$, $\det C$.

[19]We are indebted to Dr. Ling-Fong Li and Dr. M. A. Rashid for kindly showing that the solution (15) is not just an extremum of the (12 + 3)-parameter potential $V(A, B, C)$ (with three mass and twelve interaction terms) but represents a possible minimum for a specified sequence of signs of these parameters. A simple illustrative example for the sequence of signs [see Ling-Fong Li, Phys. Rev. D 9, 1723 (1973)] is provided by the (6 + 3 parameter) potential. Take

$$V(A,B,C) = \sum_{A,B,C,\alpha,\beta,\gamma} \{\alpha_1 \mathrm{Tr}[AA^\dagger] + \alpha_2 \mathrm{Tr}[(AA^\dagger)^2] + \alpha_3 \mathrm{Tr}[AA^\dagger AA^\dagger]\},$$

where (besides $\alpha_1, \beta_1, \gamma_1$) β_3 is negative, while all other parameters are positive. Arrange $\gamma_1 : \alpha_1 : \beta_1$ (the mass parameters for C, A, and B) to be in the ratios $1 : \alpha^{-2} : \alpha^{-4}$, while all other parameters are of magnitude $f^2/4\pi \approx 1$. One finds that (c_1, c_4), (a_1, a_4), and (b_4) are in the ratios $1 : \alpha^{-1} : \alpha^{-2}$ ($\alpha = \frac{1}{137}$). If $V(A, B, C)$ contains terms proportional to $\det A$, $\det B$, and $\det C$ [corresponding to the symmetry $SU(4)_L \times SU(4)_R \times SU(4')$ rather than $U(4)_L \times U(4)_R \times U(4')$], an extremum exists with

$$\langle A_D \rangle = \begin{bmatrix} a_1 & & & \\ & a_1 & & \\ & & a_3 & \\ & & & a_4 \end{bmatrix},$$

i.e., the solution exhibits a residual symmetry SU(2) × U(1). We have not examined whether this solution represents a true minimum.

[20]One might have been tempted to replace the multiplet $A = (4, \bar{4}, 1)$ of G by a more economical submultiplet $\tilde{A} = (2, 2, 1)$ of $\mathcal{G}$ with the (apparently allowed) pattern

$$\langle \tilde{A} \rangle = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix},$$

together with $\langle B \rangle$ and $\langle C \rangle$ as in (15). So far as W_L, W_R, and V masses are concerned, one might then have the desired sequence of masses (except that $W_L^{\pm} W_R^{\mp}$ mixing would be absent). However, we have preferred to use the larger multiplet A, since the global SU(3) then has a chance to emerge more naturally.

[21]By adding a new multiplet $D = (1, 1, \underline{15})$, it should be possible to make the X's superheavy without affecting the mass of S^0. Likewise, a multiplet $E = (1, \underline{3}, 1)$ would affect only the mass of $W_R^{\pm}$. With both these multiplets present, b_4 need be no larger than a_4, and S^0 may have a mass as light as $(\frac{3}{8})^{1/2} fa \propto (f/g) m_{W_L^{\pm}}$. The introduction of such multiplets does not disturb the pattern of solutions $\langle A_D \rangle$, $\langle B_D \rangle$, $\langle C_D \rangle$ shown in the text.

[22]We have retained seemingly insignificant terms of order $\Delta(g^2/f^2)$ and (c^2/b^2) in $\tilde{U}^0$ partly because the Δ terms may have observable consequences in parity violation in nuclear transitions and partly because without these terms $\tilde{U}^0$ will not appear to be anywhere near an eigenstate of the mass matrix.

[23]Note that the physical particles [with broken SU(3′)] may not be $\tilde{U}^0$ and V^0, but rather their linear combinations. This will be the case if the mass matrix assigned unequal to V_3 and V_8 before diagonalization.

[24]If we had not set the θ's and ϕ's equal to zero, all color and valency components would have mixed, and not just the strange exotic X' with W_L. An attractive choice which breaks all color symmetries is $|\theta_L - \theta_R| = 90°$.

[25]In the present scheme (with both B and C present), c_4 can be as small as we wish since the X's can be superheavy through b_4. (This is in contrast to the limited multiplet scheme exhibited in II, where only either B or C was introduced.)

[26]This contrasts with the view of D. J. Gross and F. Wilczek, Phys. Rev. D 8, 3633 (1973), and S. Weinberg, Phys. Rev. Lett. 31, 494 (1973), who have suggested from independent considerations that the $V(8)$ octet consists of massless particles and the SU(3′) color symmetry is exact; the consequent infrared infinities have been welcomed by these authors with the hope that they may prevent emission of quarks and the $V(8)$ gluons.

[27]Such a collaboration may perhaps receive a "natural" interpretation within a U(4′) structure if A and A' correspond to $(\bar{\psi}\psi)$ composites. In this case A' may play its legitimate role without spoiling SU(4′) natural symmetry and the postulate of $\langle \bar{\psi}\psi \rangle \neq 0$ (with A and A' collaborating) would provide perhaps a more attractive way of building in fermion masses.

[28]We use the words "natural symmetry" in the sense of S. Weinberg, Phys. Rev. Lett. 29, 338 (1972), and H. Georgi and S. L. Glashow, Phys. Rev. D 6, 2977 (1972). The problems of calculability (and renormalizability) of radiative corrections as well as of possible pseudo-Goldstone bosons in the present model have been considered in detail by Dr. D. Ross (see Ref. 15).

[29]Of course, if one could leave $\mathcal{O}_{a,b,c}$ quarks massless [a possibility which is attractive for chiral $SU(2)_L \times SU(2)_R$ symmetry], the $\bar{\nu}\gamma_\mu \mathcal{O} X_\mu$ interaction would be γ_5-invariant. This, however, may not apply to ν' interactions, since the companion χ quarks are presumably massive. We should remark that light quark masses (*analogous* to light lepton masses) would be an interesting possibility, if the idea of "hiding" quarks becomes feasible.

[30]By "U^0" we mean $\tilde{U}^0$ or linear combinations of U^0 and V^0 depending upon the complexion of the physical particles (see footnote 20).

[31]The possible occurrence of such states has been pointed out by M. Y. Han and Y. Nambu, Phys. Rev. 139, B1006 (1965), and has been considered subsequently by many authors.

[32]The contribution of these spin-1 gluons (which should possess an intrinsic magnetic moment) to the electromagnetic current may play a role either directly (corresponding to the production of these gluons in *pairs*) or indirectly (via the light-cone picture) in providing an explanation for the recently observed near "constancy" of $(e^+e^- \to \text{hadrons})$ over a large range of s = center-of-mass (energy)^2. The gluon pair could contribute a term growing with s in the region of interest, which might compensate for the expected $(1/s)$ fall-off of other contributions.

[33]By simple $SU(2)_L \times U(1)$ theory, we mean the gauge theory of the type suggested by Abdus Salam and J. C. Ward, Phys. Lett. 13, 168 (1964); S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967), and Abdus Salam, in *Elementary Particle Physics*, edited by N. Svartholm (Almqvist and Wiksells, Stockholm, 1968), p. 367, together with extension to hadrons as given by S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D 2, 1285 (1970).

[34]With this correspondence, the predictions of the two schemes coincide for leptonic processes provided the contributions from other exchanges (in particular S^0) are negligible. In this case, one may describe the neutral-current processes by one gauge angle ξ (see Ref. 33) given by $\sin^2\xi = g_R^2/(g_L^2 + g_R^2)$. If $g_L \propto g_R$, we expect $\sin^2\xi$ to be nearly 0.5. In line with our suggestion in Sec. II that the starting Fermi Lagrangian may be completely left-right symmetric with equality of *bare* coupling constants $g_L^{(0)}$ and $g_R^{(0)}$, the ratio of the renormalized coupling constants (g_L/g_R) would

differ from unity by calculable finite radiative corrections of order $\alpha \ln(m_{W_R}{}^2/m_{W_L}{}^2)$ to the extent that spontaneous symmetry breaking introduces different masses for W_L and W_R (for example). Thus a study of neutral-current processes and $(V+A)$ interactions could help determine whether there is a complete left-right symmetry $(L \longleftrightarrow R)$ in the basic Fermi Lagrangian.

[35]Note that strong-interaction quark-quark scattering amplitudes arising due to $V(8)$ exchange are of order $(f^2/f^2c^2) \simeq (1/c^2)$.

[36]S. Weinberg, Phys. Rev. Lett. 31, 494 (1973); R. N. Mohapatra, J. C. Pati, and P. Vinciarelli, Phys. Rev. D 8, 3652 (1973).

[37]In the classic experiments of H. S. Gurr, W. R. Kropp, F. Reines, and B. S. Meyer [Phys. Rev. 158, 1321 (1967)], attention was concentrated on relatively high-energy charged secondaries, so that possibly the two-body decays of the proton were the ones which received more emphasis. Such decays are forbidden in the model presented in this paper. One may wonder whether some of the unidentified low-energy charged secondaries could have come from four- (or more-) particle decays of the proton. In a recent paper, F. Reines and M. F. Crouch [Phys. Rev. Lett. 32, 493 (1974)] report five μ events which may possibly have resulted from proton decays in their detectors. In the authors' view, so far as this experiment is concerned, "it seems prudent to interpret the signal so as to yield a lower limit on nucleon lifetime" of 2×10^{30} years.

We are indebted to Professor F. Reines, Professor H. S. Gurr, and Professor Z. Zichichi for numerous kind discussions.

PHYSICAL REVIEW D VOLUME 17, NUMBER 1 1 JANUARY 1978

Neutral-current phenomena within the left-right-symmetric unified theory of quarks and leptons

Jogesh C. Pati* and Subhash Rajpoot
Imperial College, London, England

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

(Received 31 May 1977)

The neutral-current weak-interaction sector of the left-right-symmetric unified theory of quarks and leptons based on the symmetry structure $SU(2)_L \times SU(2)_R \times SU(4)'_{L+R}$ and its subgroup $SU(2)_L \times SU(2)_R \times SU(3)'_{L+R} \times U(1)_{L+R}$ (suggested in earlier papers) is studied in detail here. The theory admits in general of *two* weak neutral gauge bosons N_1 and N_2 [and two sets of left and right charged W's $(W_L^{\pm}, W_R^{\pm})$]. As pointed out earlier, there are two distinct possibilities for the pattern of spontaneous symmetry breaking which lead to the mass relations (i) $m_{N_1}{}^2 \sim m_{W_L^{\pm}}{}^2 \ll m_{N_2}{}^2$, (ii) $m_{N_1}{}^2 \sim m_{N_2}{}^2 \sim m_{W_L^{\pm}}{}^2$. Case (i) is identical to the familiar $SU(2) \times U(1)$ theory for all predictions. Case (ii), on the other hand, leads to departures from $SU(2) \times U(1)$. One of the crucial theorems following from our investigation is that such departures occur only for the electron-induced atomic parity experiments and *not* for neutrino-induced weak processes in the chiral γ_5-invariant limit, in which W_L-W_R mixing as well as fermion masses vanish. In this manner (*for the chiral limit defined above*), $SU(2) \times U(1)$ becomes an integral and stable ingredient of the left-right-symmetric theory for both cases (i) and (ii) insofar as their predictions for neutrino-induced reactions are concerned; the differences between the two cases (in this chiral limit) lie solely in their predictions regarding atomic parity violation and, of course, the masses of the two neutral particles N_1 and N_2. We study the chiral as well as the nonchiral cases and compare their predictions with experiments, observing that approximate or exact global chiral symmetry of the left-right-symmetric theory is a point of good agreement between theory and *all* the present neutral-current experiments. We exhibit the dependence of the atomic parity-violation parameter Q_W on the masses of the two neutral particles N_1 and N_2 and remark in particular that (depending upon the sign and magnitude of Q_W), there is the exciting experimental possibility within the left-right-symmetric theory that the mass of one of the neutral gauge particles is sensibly smaller than that of the charged $W_L^{\pm}$. This particular possibility does not materialize, if one insists on *exact* atomic parity conservation.

I. INTRODUCTION

The null or near-null result of the present atomic parity-violation experiments[1] (taken at its face value[2]) together with manifest ν-$\bar{\nu}$ distinction in neutral-current scattering experiments[3] suggests that quite possibly a radical departure from the simple $SU(2) \times U(1)$ group structure is called for even at low energies. Neither the simple $SU(2) \times U(1)$ theory[4] nor the standard vectorlike theory[5] based on the $SU(2) \times U(1)$ group can, it seems, account for the present set of experimental data.

The purpose of this note is to stress that such departures from the simple $SU(2) \times U(1)$ theory are in general well expected to occur within the *left-right-symmetric* unified theory of quarks and leptons proposed some time ago.[6] The gauge structure is *either* $\mathcal{G} = SU(2)_L \times SU(2)_R \times SU(4)'_{L+R}$, *or* its left-right-symmetric subgroup[7] $\tilde{\mathcal{G}} = SU(2)_L \times SU(2)_R \times U(1)_{L+R} \times SU(3)'_{L+R}$; both are proposed in Ref. 6. There are two key features of this theory relevant for the present discussion.

(a) To begin with, the gauge structure $SU(2)_L \times SU(2)_R$ is left-right-symmetric with left- and right-handed fermions transforming in a parallel fashion under two distinct $SU(2)$'s [$SU(2)_L$ and $SU(2)_R$]. [This is unlike both the simple $SU(2)_L \times U(1)$ theory, which treats the left- and right-handed fermions asymmetrically from the start, as well as the vectorlike $SU(2) \times U(1)$ theory for which the left- and right-handed fermions transform in a parallel manner, but under the *same* $SU(2)$ group.] In contrast, for the symmetry structure $\mathcal{G}$, for every left-handed $(V-A)$ current coupled to the gauge particles ($W_L^{\pm}$ or W_L^3), there exists a *parallel* right-handed $(V+A)$ current coupled to a distinct set of gauge particles ($W_R^{\pm}$ or W_R^3). The commuting group $SU(4)'_{L+R}$ [as well as its subgroup $SU(3)'_{L+R} \times U(1)_{L+R}$] acting on the four color indices generate just vector currents

and thus also preserve the left-right symmetry.

As a result [subject to[8] the condition that the bare gauge coupling constants $g_L^{(0)}$ and $g_R^{(0)}$ associated with the gauge groups $SU(2)_L$ and $SU(2)_R$ are equal], the gauge Lagrangian[9] of the theory prior to spontaneous symmetry breaking) is parity-conserving. Parity violation is introduced into such a theory in the charged and *independently* in the neutral-current sectors if, for example,[10] spontaneous symmetry breaking assigned *asymmetric masses* to W_L's and W_R's. The major motivation[6] for the introduction of a left-right-symmetric gauge structure was the fulfillment of the esthetic desire that parity violation may be generated spontaneously within such a theory. Additional motivations for the introduction of such a gauge structure are the elimination[6] of Abelian contributions to electric charge and the generation of a desirable milliweak theory of CP violation.[11]

(b) Second and equally important is the fact that this theory permits[6] (depending upon the pattern and parameters of spontaneous symmetry breaking) the emergence of two relatively "light" neutral gauge particles[12] (rather than just one) in addition to the photon and the color gluons. Each of these two can induce neutral-current interactions of order G_{Fermi}. Likewise the theory permits two positively charged gauge particles to remain rather light (instead of the single $W_L^{\pm}$); one of them induces the familiar charged-current interactions of strength G_F (by choice) and the other of strength $<G_F/10$ (this last restriction can be relaxed if we assume more than four quark flavors[13]). It may be noted that the availability of this novel feature involving two relatively light neutral and likewise two rather light charged gauge particles (rather than one of each variety) is a very special property of the gauge structure $\mathcal{G} = SU(2)_L \times SU(2)_R \times SU(4)_{L+R}$ (as well as of its left-right-symmetric subgroup[7] $\bar{\mathcal{G}}$ and all *its* extensions to superunifying groups G); it is not available, by contrast within unification hypotheses such as the ones based on the group structures[14] SU(5) and E_7. The availability of this new feature within the group structure $\mathcal{G}$ was noted in Ref. 6, where the appropriate choice of Higgs-Kibble multiplets permitting the emergence of the said solution was also given.

To be specific, it was pointed out in Ref. 6 that depending upon the choice of Higgs-Kibble multiplets and the pattern and magnitudes of their vacuum-expectation values, there are basically two extreme alternative modes[15] by which the symmetry $\mathcal{G}$ (or its superunifying extensions G) could descend spontaneously into the level of the symmetry ($\mathcal{G}_{\text{obs}}$) of the known interactions. These are schematically presented in Fig. 1. Here $U(1)_f$ denotes the Abelian symmetry generated by the color-singlet part of the electromagnetic charge (which is the entire electromagnetic charge if quarks are fractionally charged). For the case of integer-charge quarks, the symmetry $U(1)_f \times SU(3)'_{L+R}$ breaks down further[6] by a mass scale of order 1 GeV to the local symmetry generated by electric charge (which in this case becomes a sum of the color-singlet piece [$U(1)_f$] and an appropriate combination of color-octet pieces). For the present discussions it is immaterial, however, whether we choose fractional or integer-charges for the quarks (see later).

In Ref. 6, only the first case (case I) was exhibited in some detail for the sake of simplicity, although the basic ingredients needed to generate the second alternative as well as its general significance for neutral-current interactions due to the presence of *two* light neutral gauge particles [as opposed to one for the $SU(2) \times U(1)$ gauge structure] were spelled out. In this paper, we examine this second alternative in some detail and present the variety of new and experimentally interesting complexions, which could arise within this alternative by permitting the full generality of symmetry breaking, proposed in Ref. 6. We remark that atomic parity violation can be arbitrarily small within a *class of solutions* within the second alternative, while the distinction between neutrinos and antineutrinos even in neutral-current scattering experiments can be large. This last feature arises simply from the fact that the available neutrinos are produced only via parity-violating charged-current interactions and carry definite handedness (see later).

Before proceeding further, it needs to be noted that two sets of authors (Mohapatra and Sidhu,[16] and De Rújula, Georgi, and Glashow[17]) have recently carried out investigations in this area. Our work and results partly overlap with the work of these authors to the extent that they are examining the consequences of the same group structure $\bar{\mathcal{G}} = SU(2)_L \times SU(2)_R \times U(1) \times SU(3)'_{L+R}$, which was proposed in Ref. 6. [De Rújula *et al.* differ from Ref. 6 in their choice of fermion representation and somewhat also in the scheme of spontaneous symmetry breaking. On the other hand, the fermion representation with four flavors considered by Mohapatra and Sidhu (MS) is the one suggested in Ref. 6; also, the scheme of spontaneous symmetry breaking considered by MS is happily in *one-to-one-correspondence*[18] with the one proposed in Ref. 6 (see later).] We, however, examine the problem in this note from the standpoint of the breaking of the bigger symmetry via the chain $\mathcal{G} \to \bar{\mathcal{G}} \to \mathcal{G}_{\text{obs}}$.

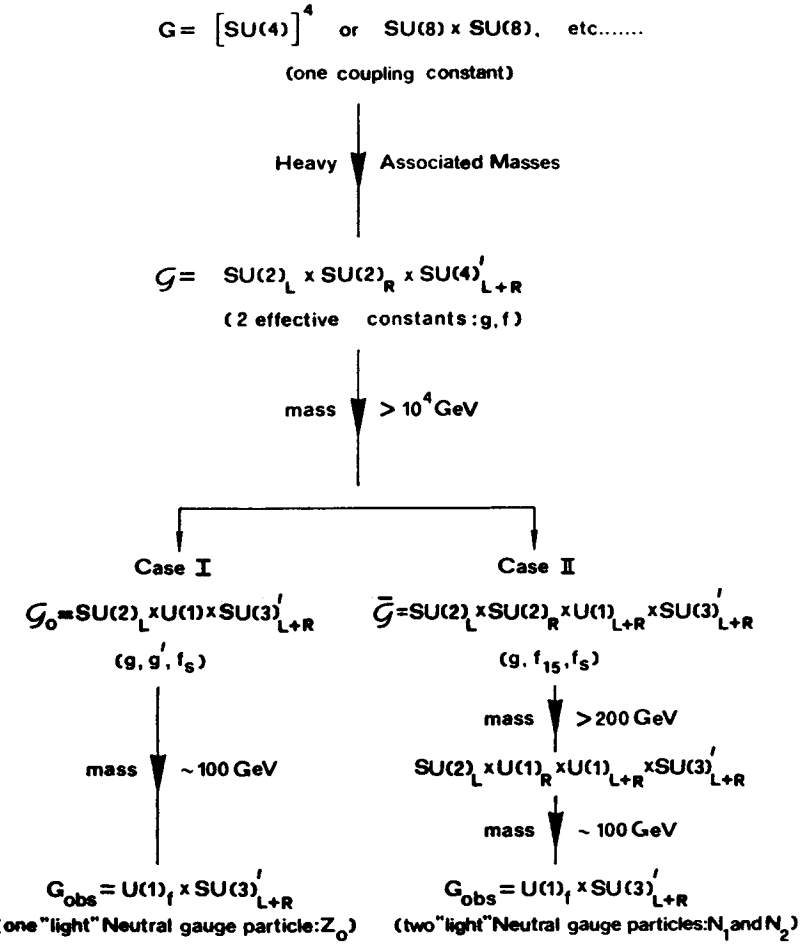


FIG. 1. Hierarchical breakdown of the left-right–symmetric unifying symmetry G. The parameters (g, g', f_s) and (g, f_{15}, f_s) denote effective coupling constants for the gauge groups $\mathcal{G}_0$ and $\bar{\mathcal{G}}$, respectively, their low-energy values depending upon the choice of the symmetry G as well as on the pattern and parameters of its spontaneous breakdown. Alternative patterns of symmetry breakdown other than those exhibited above are permissible (Ref. 15). The general considerations of neutral-current phenomena for *all* these alternatives are however the *same* as those for case I and case II.

Thus, we remark that the solutions of vanishing[16] or small[17] atomic parity violation presented in the above references are only *special cases* within a general class of solutions proposed in Ref. 6 for the breakdown of the symmetry structure $\mathcal{G}$. All solutions belonging to this class share the common feature of vanishing or small atomic parity violation, but they differ from each other notably in the predictions of the masses of the neutral particles N_1 and N_2. As we shall show, within this class of solutions there exists the distinct possibility that one of the neutral masses can be sensibly lighter than the charged W^+. This particular feature is not obtainable either for the case of exact atomic parity conservation considered by MS or for the special type of spontaneous symmetry breaking considered by De Rújula *et al.*[17]

One of the most interesting (and somewhat unexpected) theorems that we prove is that neutral-current processes initiated by left-handed neutrinos (or right-handed antineutrinos) are independent of the expectation values of those scalar multiplets which do not couple to the W_L's, provided W_L and W_R mixing or equivalently quark-mass terms are ignored. *This has the remarkable consequence that the theory (subject to the neglect of W_L-W_R mixing) agrees in all its predictions at zero momentum transfer with those*

of the familiar SU(2)×U(1) theory for neutral-current neutrino processes, while it differs from the latter in its predictions for atomic parity violation and the masses of N_1 and N_2.

The class of solutions permitting two relatively light neutral gauge particles is described in general in terms of three parameters (essentially two for most purposes; see Sec. III), which represent the masses of the two relatively light neutral particles and one coupling constant to be contrasted with two parameters *(in general)* for the SU(2)×U(1) theory.

We obtain a good description of the data involving at least eight independent neutral-current parameters: $R_\nu, R_{\bar\nu}, C_{\nu_\mu e}, C_{\bar\nu_\mu e}, C_{\bar\nu_e e}, Q_W$ (atomic), and two parameters describing elastic νp and $\bar\nu p$ scatterings. The left-right-symmetric gauge structure[6] can thus be well tested once the body of neutral-current parameters settles experimentally.

II. SALIENT FEATURES OF THE BASIC MODEL; SPONTANEOUS SYMMETRY BREAKING AND GAUGE MASSES

For the sake of completeness we wish to recall the following salient features of the basic model[6]:

(a) The model introduces a sixteen fold of *basic* four-component fermions with four flavors ($\mathcal{P}, \mathfrak{N}$, λ, c) and four colors (red, yellow, blue, and lilac), the fourth color being identified with lepton number $L \equiv L_e + L_\mu$:

$$F_{L,R} = \begin{bmatrix} \mathcal{P}_r & \mathcal{P}_y & \mathcal{P}_b & \mathcal{P}_l = \nu_e \\ \mathfrak{N}_r & \mathfrak{N}_y & \mathfrak{N}_b & \mathfrak{N}_l = e^- \\ \lambda_r & \lambda_y & \lambda_b & \lambda_l = \mu^- \\ c_r & c_y & c_b & c_l = \nu_\mu \end{bmatrix}_{L,R} . \tag{1}$$

Note the necessity for the left-right-symmetric right-handed partners (ν_{eR} and $\nu_{\mu R}$) of the left-handed neutrinos (ν_{eL} and $\nu_{\mu L}$).

(b) To implement left-right discrete symmetry as well as quark-lepton unification, the model assumes a minimal local symmetry

$$\mathcal{G} = \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \times \mathrm{SU}(4)'_{L+R} , \tag{2}$$

where $\binom{\mathcal{P}}{\mathfrak{N}}_{L,R}$ and $\binom{c}{\lambda}_{L,R}$ *for each of the four colors* (this includes the leptons) are treated as $\mathrm{SU}(2)_{L,R}$ doublets, while $(q_r, q_y, q_b, q_l)_{L+R}$ *for each of the four flavors* are treated as SU(4)′ antiquartets ($\bar 4$'s) (this puts quarks and leptons into one multiplet).

Following the suggestions in Ref. 6 and 8, we impose a discrete left-right symmetry on the theory; thus the renormalized gauge-coupling constants g_L and g_R associated with the gauge groups $\mathrm{SU}(2)_L$ and $\mathrm{SU}(2)_R$ are equal up to finite order-α corrections $[(g_L/g_R) = 1 + O(\alpha)]$. We will set throughout $g_L = g_R \equiv g$.

The gauge particles of the theory are conveniently represented by the matrices:

$$W_{L,R} = \frac{1}{2}\begin{pmatrix} \vec\tau\cdot\vec W & 0 \\ 0 & \tau_1(\vec\tau\cdot\vec W)\tau_1 \end{pmatrix}_{L,R} , \tag{3}$$

$$V = \frac{1}{\sqrt 2}\begin{pmatrix} V(8) - \dfrac{S^0\times 1}{\sqrt{12}} & X \\ \bar X & \dfrac{3}{4}^{1/2} S^0 \end{pmatrix} , \tag{4}$$

where $V(8)$ denotes the SU(3)′ octet of color gluons, X is an exotic SU(3)′ triplet ($B = -L = 1$), and S^0 is an SU(3)′ singlet associated with the 15th generator of SU(4)′.

Defining $\nabla_\mu F = \partial_\mu F' + igW_\mu F - ifFV_\mu$, where f denotes the gauge coupling constant associated with the SU(4)′ group,[19] the Fermi Lagrangian of the theory is given by

$$\mathcal{L}_{\text{Fermi}} = -\mathrm{Tr}\,\bar F_L(\gamma_\mu\nabla_\mu)_L F_L + (L\to R) , \tag{5}$$

where $F_{L,R} = \frac{1}{2}(1\pm\gamma_5)F$. *Note the manifest completely left-right-symmetric parity conserving nature of the gauge interactions.*

(c) The next feature of the model is spontaneous symmetry breaking. The model[6] introduces the following sets of Higgs-Kibble multiplets to induce a *general pattern* of spontaneous symmetry breaking (see Ref. 6, Sec. IV and footnote 21):

$$A = (2+2, 2+2, 1)\quad B = (1, 2+2, \bar 4) ,$$
$$C = (2+2, 1, \bar 4),\quad D = (1, 1, 15). \tag{6}$$

Note that the choice of the multiplets A, B, C, D is in accordance with the requirements of left-right symmetry with B balancing C. If one wishes to assign a heavy mass to the charged $W_R^\pm$ *without* at the same time making the charged $W_L^\pm$ and the neutral components $W^3_{L,R}$ and S_0^0 heavy likewise, one must introduce[6] an additional multiplet $E_R = (1, 3, 1)$ and therefore, following left-right symmetry, one must introduce a parallel multiplet[8] $E_L = (3, 1, 1)$. (The introduction of E_L would be inconsequential operationally in the sense that we will assume for simplicity[20] that $\langle E_L\rangle$ is zero or small [perhaps $O(\alpha)$] compared to $\langle A\rangle$ and/or $\langle C\rangle$.) Additional scalar multiplets may also be present to permit the emergence of fermion masses appropriately[21]; however, we assume (as in Ref. 6) that such multiplets have small vacuum expectation values and are inconsequential insofar as gauge masses are concerned.

Following Ref. 6 one may choose the parameters of the Higgs potential (lying in a specified range) so as to obtain the following sets of vacuum expectation values consistent with minimization of the potential.

$$\langle A\rangle = \begin{bmatrix} a_1 & & & \\ & a_1 & & \\ & & a_1 & \\ & & & a_4 \end{bmatrix},$$

$$\langle B\rangle = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & b_4 \end{bmatrix},$$

$$\langle C\rangle = \begin{bmatrix} c_1 & & & \\ & c_1 & & \\ & & c_1 & \\ & & & c_4 \end{bmatrix},$$

$$\langle D\rangle = \begin{bmatrix} d & & & \\ & d & & \\ & & d & \\ & & & -3d \end{bmatrix}, \qquad (7)$$

$$\langle E_R\rangle = \begin{bmatrix} 0 \\ \epsilon_R \\ 0 \end{bmatrix},$$

$$\langle E_L\rangle = \begin{bmatrix} 0 \\ \epsilon_L \\ 0 \end{bmatrix} \approx 0.$$

We have not exhibited the Cabibbo-rotated forms for the vacuum expectation matrices, since once again these are inconsequential for gauge masses (see Ref. 6 for details). The above pattern can be preserved,[6,8] consistent with the demands of natural left-right symmetry [i.e., maintenance of the choice[6] $g_L^{(0)} = g_R^{(0)}$].

Integer versus fractional quark charges. Before proceeding further, it should be emphasized that the unification hypothesis based on the symmetry structure 9 permits[6] quark charges to be *either* integral ($c_1 \neq 0$) *or* fractional ($c_1 = 0$).

$$c_1 = \text{finite} \sim 1 \text{ GeV} \qquad (8a)$$

(quarks acquire integer charges; only photon remains massless; the octet of color gluons acquire finite mass),

$$c_1 = 0 \qquad (8b)$$

(quarks acquire fractional charges; photon and the *octet* of color gluons remain massless).

Whether c_1 is finite or zero thus makes a world of difference, in general. However, as we are interested in this work in studying neutral-current phenomena *below threshold for color production* and also because the parameter c_1 (even when it is nonvanishing) is still negligible compared to the other relevant mass scales in the problem, which are of order 100 GeV, it is fortunately *immaterial* for our present considerations whether we set[22] $c_1 = 0$ or $c_1 = \text{finite} \sim 1$ GeV. *Thus for simplicity of writing we henceforth set $c_1 = 0$ bearing in mind, however, that all our considerations in the present work regarding neutral- and charged-current interactions below threshold for color production are equally valid for the cases of integer and fractionally charged quarks.* (A full analysis involving $c_1 \neq 0$ and its implications on leptoproduction of color above color threshold in light of the present work will be presented elsewhere.)

Setting $c_1 = 0$, we may now simply read off the gauge mass matrices from Eq. (17) and footnote 21 of Ref. 6. These are given below.

$$m_{X,\bar{X}}^2 = \tfrac{1}{2} f_X^2 (b^2 + c^2 + d^2), \qquad (9a)$$

$$m_{(W_L^{\pm}, W_R^{\pm})}^2 = \frac{g^2}{4} \begin{pmatrix} (a^2 + c^2 + \epsilon_L^2) & 2(a_1^2 + a_1 a_4) \\ 2(a_1^2 + a_1 a_4) & (a^2 + b^2 + \epsilon_R^2) \end{pmatrix}, \qquad (9b)$$

$$m_{\text{neutral}}^2 = \frac{1}{4} \begin{bmatrix} & W_L^3 & W_R^3 & S^0 \\ & g^2(a^2 + c^2) & -g^2 a^2 & \sqrt{3/2} f_{15} g c^2 \\ & -g^2 a^2 & g^2(a^2 + b^2) & \sqrt{3/2} f_{15} g b^2 \\ & \sqrt{3/2} f_{15} g c^2 & \sqrt{3/2} f_{15} g b^2 & (3/2) f_{15}^2 (b^2 + c^2) \end{bmatrix}, \qquad (9c)$$

where $a^2 \equiv 3a_1^2 + a_4^2$, $b^2 \equiv b_4^2$, and $c^2 \equiv c_4^2$. We have inserted f_{15} and f_s (rather than f) in the entries corresponding to the gauge particles $S_0 \equiv V_{15}$ and X, respectively, to denote the fact that the effective renormalized coupling constants associated with these gauge particles are in general expected to be quite different from that associated with the octet of color gluons owing to differing finite renormalization effects, a feature which was suggested in general terms by two of us.[23] This has recently most strongly been emphasized by Elias and Shafi,[24] who show that $f_s^2 \gg f_{15}^2$, where f_s denotes the effective quark-gluon coupling constant. We shall incorporate these results within our analysis in this note.

Note the important feature that the multiplet $D=(1,1,15)$ provides mass *only* to the exotic gauge particles X and $\bar{X}$, which are coupled to $\bar{q}\gamma_\mu l$ and $\bar{l}\gamma_\mu q$ currents, respectively, but not to $V(8)$ and S^0. Thus D splits SU(4)′ naturally to $SU(3)'_{L+R} \times U(1)_{L+R}$, where the U(1) corresponds to the 15th generator of the SU(4)′ group; this generator is associated with the gauge particle $S^0 \equiv V_{15}$. Likewise the multiplets $E_R=(1,3,1)$ and $E_L=(3,1,1)$ provide masses only to the charged gauge particles $W_R^\pm$ and $W_L^\pm$, respectively, but *not* to the neutral ones.

Thus note the interesting feature of the model[6]: it is possible to destroy the left-right symmetry altogether in the mass matrix of the charged gauge bosons and thereby in charged-current interactions (by choosing, e.g., $\epsilon_R \gg \epsilon_L, a, b, c$) and *yet* preserve it approximately or exactly in the mass matrix of the neutral-gauge bosons and thereby in the neutral-current interactions by choosing $b \approx c$ or $b = c$ [see Eqs. (9b) and (9c)]. Vice versa is however not possible. The full implications of this remark for neutral-current experiments will be apparent in a moment.

Phenomenologically, from the known degree of forbiddenness of $K_L - \bar{\mu}e$, one may deduce that

$$m_X > 10^4 \text{ GeV} \tag{10}$$

for the basic model. Likewise from the known suppression of $(V+A)$ interactions in β decay, one may deduce[6] (within the four-flavor model[25]) that

$$m_{W_R^+} > 3\left(\frac{\cos\theta_R}{\cos\theta_L}\right)^{1/2} m_{W_L^+}, \tag{11}$$

where θ_L and θ_R denote the Cabibbo angles for the left- and right-handed fermions.[26]

We are now ready to point out how, in accordance with the above phenomenological constraints, cases I and II, representing the two extreme alternative modes for the breakdown of $\mathcal{G}$ to $\mathcal{G}_{obs}$, arise (these two alternative cases are described in Sec. I).

Case I: This case arises if in the first place we choose the vacuum expectation value of the "right" multiplet $B=(1,2+2,\bar{4})$ to be much larger than those of the "left" multiplet $C=(2+2,1,\bar{4})$, and the multiplet $A=(2+2,2+2,1)$, i.e.,

$$b \gg c \text{ and } a.$$

Such a choice, we note, destroys left-right symmetry in the charged- *as well as* in the neutral-current sectors, since with this choice all three right-handed gauge particles ($W_R^\pm, W_R^3$) acquire heavier masses than the corresponding left-handed ones ($W_L^\pm, W_L^3$) [see Eqs. (9b) and (9c)].

As an extreme possibility, if we furthermore choose b to be the major contributor to the X-mass (i.e., $b > 10^4$ GeV), then left-right symmetry as well as quark-lepton unification are lost in *one* stroke through $\langle B \rangle$; the symmetry $\mathcal{G}$ thereby reduces[27] to $\mathcal{G}_0 = SU(2)_L \times U(1) \times SU(3)'_{L+R}$. In turn, $\mathcal{G}_0$ breaks through $\langle C \rangle$ and $\langle A \rangle \neq 0$ down to $\mathcal{G}_{obs}$ generating just one light neutral gauge particle, the Z^0. (Note that for this case the presence of the multiplets D, E_L, and E_R is inessential for a gross view, since *both* X and $W_R^\pm$ become heavy or superheavy through $\langle B \rangle$.)

Case II. This case arises[6] provided we introduce the multiplets like $D=(1,1,15)$, $E_R=(1,3,1)$, and $E_L=(3,1,1)$ in addition to $\langle A \rangle$, $\langle B \rangle$, and $\langle C \rangle$ with the choice of $d > 10^4$ GeV $\gg \epsilon_{R,L}, a, b, c$. In this case, the theory loses quark-lepton unification first [i.e., $SU(4)'_{L+R} \to U(1)_{L+R} \times SU(3)'_{L+R}$] since $\langle D \rangle$ is large, but it still retains its intrinsic left-right–symmetric nature; the symmetry $\mathcal{G}$ reduces to the left-right–symmetric structure

$$\bar{\mathcal{G}} = SU(2)_L \times SU(2)_R \times U(1)_{L+R} \times SU(3)'_{L+R}.$$

The manner in which $\bar{\mathcal{G}}$ reduces farther down to $\mathcal{G}_{obs} = U(1)_f \times SU(3)'_{L+R}$ depends very much on the relative magnitude of the parameters $\epsilon_{R,L}$ and a, b, c (for the neutral-current sector only the three parameters a, b, c are relevant).

If one allows $\epsilon_R \gg \epsilon_L, a, b, c$ (so that $m_{W_R^+} > 3m_{W_L^+}$), one loses left-right symmetry in the *charged-current* sector, leading to predominance of $V-A$ over $V+A$ in this sector. One may still retain left-right symmetry in the neutral-current sector, however, as long as $b = c$, which ensures that the mass parameters associated with W_L^3 and W_R^3 are equal [see Eq. (9c)]. *Thus departures from the left-right symmetry in the neutral-current sector for the left-right–symmetric theory[6] are strictly proportional to* $(b^2 - c^2)/(b^2 + c^2)$.

Atomic parity violation would clearly be small if $b \sim c$, vanishing[28] in the limit $b = c$. In *either* case ($b \sim c$ or $b = c$), one expects[6] to see *two* relatively light neutral gauge particles N_1 and N_2. The case $b \sim c$ (rather than $b = c$) allows the inter-

esting possibility that one of the neutral gauge particles can be sensibly lighter than the charged W_L^+ with the magnitude of atomic parity violation still remaining small (by about a factor ≈ 2 to 4) compared to that of the SU(2) × U(1) theory. We discuss this case as well as the case $b = c$ and their interrelationship in some detail in the next two sections.

III. GENERAL COMPOSITIONS OF THE NEUTRAL GAUGE PARTICLES AND THEIR INTERACTIONS WITHIN THE LEFT-RIGHT-SYMMETRIC GAUGE STRUCTURE $\mathcal{G}$ OR $\overline{\mathcal{G}}$

First we note the composition of the massless photon. This, for the two alternative cases of fractional and integer charge quarks ($c_1 = 0$ and $c_1 \neq 0$) is given by[6]

$$A_\mu = e\left[\frac{W_L^3 + W_R^3}{g} - \frac{\sqrt{2/3}\,S^0}{f_{15}}\right] \tag{12a}$$

(fractionally charged quarks; $c_1 = 0$),

$$A_\mu = e\left[\frac{W_L^3 + W_R^3}{g} - \frac{\sqrt{2/3}\,S^0}{f_{15}} + \frac{(2/\sqrt{3})U^0}{f_s}\right] \tag{12b}$$

(integer charged quarks; $c_1 \neq 0$). Here $U^0 \equiv \frac{1}{2}(\sqrt{3}\,V_3 + V_8)$; f_s and f_{15} denote the effective renormalized gauge coupling constants associated with the color-gluon U^0 and the gauge particle $V_{15} \equiv S^0$, respectively; the electric charge for the two cases is simply related to these gauge coupling constants as follows:

$$\frac{1}{e^2} = \frac{2}{g^2} + \frac{2}{3}\frac{1}{f_{15}^2} \quad \text{(f.c.q.)}, \tag{13a}$$

$$\frac{1}{e^2} = \frac{2}{g^2} + \frac{2}{3}\frac{1}{f_{15}^2} + \frac{4}{3}\frac{1}{f_s^2} \quad \text{(i.c.q.)}. \tag{13b}$$

The abbreviations f.c.q. and i.c.q. stand for fractionally charged and integer charged quarks, respectively. Since we expect $f_s^2 \gg f_{15}^2 \sim g^2$ owing to differing finite renormalization effects,[23,24] we see that either of the above two equations [i.e., (13a) or (13b)] leads essentially to the same value of g^2 for a given ratio (f_{15}^2/g^2). This is yet another reason why there is no difference worth noting between the cases of fractional and integer charge quarks insofar as neutral-current interactions *below* threshold for color production are concerned.

From here on, therefore, we substitute as mentioned in Sec. II, $c_1 = 0$ for simplicity bearing in mind that our considerations in this note, involving neutral-current interactions below color threshold, hold equally for both cases: $c_1 = 0$ (f.c.q.) and $c_1 \neq 0$ (i.c.q.).

The photon [Eq. (12a)] and the two other eigenstates (N_1 and N_2) of the neutral mass matrix (9c) are then given by

$$A_\mu = \frac{W_L^3 + W_R^3}{\sqrt{2}}\cos\alpha - S^0\sin\alpha, \tag{14a}$$

$$N_1 = Y_1\cos\beta + Y_2\sin\beta, \tag{14b}$$

$$N_2 = -Y_1\sin\beta + Y_2\cos\beta, \tag{14c}$$

where

$$Y_1 \equiv \frac{W_L^3 - W_R^3}{\sqrt{2}}, \tag{15}$$

$$Y_2 \equiv \frac{W_L^3 + W_R^3}{\sqrt{2}}\sin\alpha + S^0\cos\alpha, \tag{16}$$

$$\cos\alpha = \frac{\sqrt{2}e}{g}, \quad \sin\alpha = \frac{\sqrt{2/3}e}{f_{15}}. \tag{17}$$

Note that the normalization condition (i.e., $\cos^2\alpha + \sin^2\alpha = 1$) is the same as Eq. (13a) (corresponding to the case $c_1 = 0$). The mixing angle β, which determines the composition of the physical particles N_1 and N_2 as well as their masses, may conveniently be obtained from (9c) by first transforming the same into the orthonormal basis (A_μ, Y_1, and Y_2). The photon, being an eigenstate, decouples from Y_1 and Y_2; the mass matrix of the $(Y_1$-$Y_2)$ complex, thus obtained, is given by

$$m_{(Y_1,Y_2)}^2 = \begin{array}{c} \\ Y_1 \\ Y_2 \end{array}\begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_2 \end{array} \\ \begin{bmatrix} \frac{g^2}{8}(4a^2 + b^2 + c^2) & -\frac{\sqrt{3}g}{8}\left(f_{15}^2 + \frac{g^2}{3}\right)^{1/2}(b^2 - c^2) \\ -\frac{\sqrt{3}g}{8}\left(f_{15}^2 + \frac{g^2}{3}\right)^{1/2}(b^2 - c^2) & \frac{3}{8}\left(f_{15}^2 + \frac{g^2}{3}\right)(b^2 + c^2) \end{bmatrix} \end{array}. \tag{18}$$

Note in the special limit $b = c$, Y_1 and Y_2 are the eigenstates of the mass matrix, as in this limit they are decoupled from each other; their masses are given simply by the diagonal elements of the mass matrix (18). For the general situation, $b \neq c$, the physical particles are N_1 and N_2 [Eq. (14)]. Diagonalizing the matrix (18), we obtain

$$m_{N_i}^2 = \tfrac{1}{16}(f_{15}^2 + \tfrac{2}{3}g^2)(b^2 + c^2) + \tfrac{1}{48}g^2a^2 \mp \left\{\left[\tfrac{3}{16}f_{15}^2(b^2 + c^2) - \tfrac{1}{48}g^2a^2\right]^2 + \tfrac{3}{64}g^2(f_{15}^2 + \tfrac{1}{3}g^2)(b^2 - c^2)^2\right\}^{1/2}, \tag{19a}$$

$$\tan 2\beta = \frac{\sqrt{3}g(f_{15}^2 + g^2/3)^{1/2}(b^2 - c^2)}{\frac{3}{2}f_{15}^2(b^2 + c^2) - 2g^2a^2}, \tag{19b}$$

where, in Eq. (19a), the upper (lower) sign refers to N_1 (N_2). The neutral-current interaction induced by N_1 and N_2 is given by

$$\mathcal{L}_{NC} = (J_\mu^{Y_1}\cos\beta + J_\mu^{Y_2}\sin\beta)N_{1\mu} + (-J_\mu^{Y_1}\sin\beta + J_\mu^{Y_2}\cos\beta)N_{2\mu}, \tag{20}$$

$$J_\mu^{Y_1} = \frac{g}{2\sqrt{2}}[(\nu_{eL}\gamma_\mu\nu_{eL} - \bar{\nu}_{eR}\gamma_\mu\nu_{eR} + \bar{\nu}_{\mu L}\gamma_\mu\nu_{\mu L} - \bar{\nu}_{\mu R}\gamma_\mu\nu_{\mu R}) - (\bar{e}\gamma_\mu\gamma_5 e + \bar{\mu}\gamma_\mu\gamma_5\mu)$$
$$+ (\bar{\mathcal{P}}\gamma_\mu\gamma_5\mathcal{P} + \bar{c}\gamma_\mu\gamma_5 c - \bar{\mathcal{R}}\gamma_\mu\gamma_5\mathcal{R} - \bar{\lambda}\gamma_\mu\gamma_5\lambda)], \tag{21}$$

$$J_\mu^{Y_2} = \frac{g}{2\sqrt{2}(3r+1)^{1/2}}[(1-r)(\bar{\mathcal{P}}\gamma_\mu\mathcal{P} + \bar{c}\gamma_\mu c) - (1+r)(\bar{\mathcal{R}}\gamma_\mu\mathcal{R} + \bar{\lambda}\gamma_\mu\lambda)$$
$$+ (3r+1)(\bar{\nu}_{eL}\gamma_\mu\nu_{eL} + \bar{\nu}_{eR}\gamma_\mu\nu_{eR} + \bar{\nu}_{\mu L}\gamma_\mu\nu_{\mu L} + \bar{\nu}_{\mu R}\gamma_\mu\nu_{\mu R}) + (3r-1)(\bar{e}\gamma_\mu e + \bar{\mu}\gamma_\mu\mu)], \tag{22}$$

$$r \equiv (f_{15}^2/g^2). \tag{23}$$

Note that $J_\mu^{Y_1}$ is a pure axial-vector current for all four-component fields (this applies to neutrinos as well, if neutrinos are light four-component objects) and $J_\mu^{Y_2}$ is a pure vector current. Thus *in the limit* $b = c$ ($\tan 2\beta = 0$), *for which* N_1 *and* N_2 *reduce to* Y_1 *and* Y_2, *respectively, the neutral-current interaction is entirely parity conserving* (*for electrons, quarks, and neutrinos as well*) as mentioned before. There would still be ν-$\bar{\nu}$ distinction in neutral-current scattering since the neutrinos and antineutrinos are produced by parity-violating charged-current interactions, which yield ν and $\bar{\nu}$'s with definite handedness (i.e., neutrinos with left helicities and antineutrinos with right helicities). Such neutrinos *and* antineutrinos contribute to neutral-current scattering only through the term $\bar{\nu}_L\gamma_\mu\nu_L$ in the currents $J_\mu^{Y_{1,2}}$, but not through the term $\bar{\nu}_R\gamma_\mu\nu_R$ to the extent that the mass of the neutrino may be neglected. The current $\bar{\nu}_L\gamma_\mu\nu_L$ is $V-A$ for neutrinos, but $V+A$ for antineutrinos, hence the distinction between ν and $\bar{\nu}$ in neutral-current scattering *even* for the intrinsically parity-conserving Lagrangian ($b = c$). Note that unlike neutrinos, electrons and quarks appear with both helicities for processes of interest; thus in the limit $b = c$ atomic interactions will not show any parity violation.

Before plunging into phenomenology we must note the charged-current constraint, i.e.,

$$g^2/8m_{W_L^+}^2 = G_F/\sqrt{2}. \tag{24}$$

For a given $r \equiv f_{15}^2/g^2$, we thus obtain by using Eqs. (13a) and (24)

$$m_{W_L^+}^2 = \left(1 + \frac{1}{3r}\right)\frac{e^2}{\sqrt{8}}G_F^{-1} \approx \left(1 + \frac{1}{3r}\right)(53\ \text{GeV})^2. \tag{25}$$

Furthermore, neglecting the small perturbation of W_L^+-W_R^+ mixing (compared to the mass of W_R^+) as well as[29] ϵ_L in (9b), we also have

$$m_{W_L^+}^2 \approx \tfrac{1}{4}g^2(a^2 + c^2). \tag{26}$$

Combining Eqs. (24) and (25), we derive

$$a^2 + c^2 = \frac{1}{\sqrt{2}G_F} = \frac{10^5 m_N^2}{\sqrt{2}}. \tag{27}$$

Thus, of the three parameters a^2, b^2, and c^2 only two are independent.

Correspondingly, we define two parameters[30]

$$\zeta \equiv \frac{c^2}{c^2 + a^2}; \quad \eta \equiv \frac{b^2}{c^2 + a^2}. \tag{28}$$

Thus,

$$0 \leq \zeta \leq 1, \quad 0 \leq \eta < \infty. \tag{29}$$

It is convenient to express the masses of N_1 and N_2, the mixing angle β, and the amplitudes for the various neutral-current processes in terms of $r \equiv f_{15}^2/g^2$, ζ and η; these we list below

$$K_1 \equiv \frac{m_{W_L^+}^2}{m_{N_1}^2} = \frac{4}{P_1 - P_2}, \tag{30a}$$

$$K_2 \equiv \frac{m_{W_L^+}^2}{m_{N_2}^2} = \frac{4}{P_1 + P_2}, \tag{30b}$$

$$\tan 2\beta = \frac{(3r+1)^{1/2}(\eta - \zeta)}{\frac{3}{2}r(\zeta + \eta) - 2(1 - \zeta)}, \tag{31}$$

$$P_1 \equiv (3r+2)(\zeta + \eta) + 4(1 - \zeta), \tag{32}$$

$$P_2 \equiv \{[3r(\zeta + \eta) - 4(1 - \zeta)]^2 + 4(3r+1)(\zeta - \eta)^2\}^{1/2}. \tag{33}$$

The amplitudes. We define the relevant neutral-current amplitudes by the following effective Lagrangian (here $\mathcal{P}$ and $\mathcal{R}$ denote quark fields):

$$\mathcal{L}_{\rm NC}^{\rm eff}=\frac{G_F}{\sqrt{2}}[\bar\nu_\mu\gamma_\mu(1+\gamma_5)\nu_\mu][\bar{\mathcal{P}}\gamma_\mu(g_V^{\mathcal{P}}+g_A^{\mathcal{P}}\gamma_5)\mathcal{P}+\bar{\mathcal{N}}\gamma_\mu(g_V^{\mathcal{N}}+g_A^{\mathcal{N}}\gamma_5)\mathcal{N}+\bar e\gamma_\mu(M_V^e+M_A^e\gamma_5)e+\cdots]$$
$$+\frac{G_F}{\sqrt{2}}[\bar\nu_e\gamma_\mu(1+\gamma_5)\nu_e]\bar e\gamma_\mu(E_V^e+E_A^e\gamma_5)e+\frac{G_F}{\sqrt{2}}(\bar e\gamma_\mu\gamma_5 e)(C_V^{\mathcal{P}}\bar{\mathcal{P}}\gamma_\mu\mathcal{P}+C_V^{\mathcal{N}}\bar{\mathcal{N}}\gamma_\mu\mathcal{N})+\cdots . \tag{34}$$

Defining two special combinations

$$G_V\equiv\left[K_1\sin^2\beta+K_2\cos^2\beta+\frac{(K_1-K_2)\sin\beta\cos\beta}{(3\gamma+1)^{1/2}}\right], \tag{35}$$

$$G_A\equiv[K_1\cos^2\beta+K_2\sin^2\beta+(K_1-K_2)(3\gamma+1)^{1/2}(\sin\beta\cos\beta)], \tag{36}$$

which in terms of ζ and η acquire the simple forms

$$G_V=\frac{\eta+2(1-\zeta)}{(3\gamma+1)(\zeta+\eta-\zeta^2)}, \tag{37}$$

$$G_A=\frac{\eta}{(\zeta+\eta-\zeta^2)}, \tag{38}$$

we now show that the amplitudes of *all* neutrino-induced processes as defined by (34) are simply given in terms of either G_V or G_A, though not the the electron-induced process.

Process

$$\nu\mathcal{P}\to\nu\mathcal{P}\begin{cases}g_V^{\mathcal{P}}=(\gamma-1)G_V\\ g_A^{\mathcal{P}}=-G_A\end{cases}$$

$$\nu\mathcal{N}\to\nu\mathcal{N}\begin{cases}g_V^{\mathcal{N}}=(\gamma+1)G_V\\ g_A^{\mathcal{N}}=G_A\end{cases}$$

$$\nu_\mu e\to\nu_\mu e\begin{cases}M_V^e=-(3\gamma-1)G_V\\ M_A^e=G_A\end{cases}$$

$$\nu_e e\to\nu_e e\begin{cases}E_V^e=-1-(3\gamma-1)G_V\\ E_A^e=-1+G_A\end{cases} \tag{39}$$

$$e\mathcal{P}\to e\mathcal{P}\;\{C_V^{\mathcal{P}}=-\frac{(\gamma-1)(K_1-K_2)\sin2\beta}{2(3\gamma+1)^{1/2}}$$

$$e\mathcal{N}\to e\mathcal{N}\;\{C_V^{\mathcal{N}}=-\frac{(\gamma+1)(K_1-K_2)\sin2\beta}{2(3\gamma+1)^{1/2}}.$$

Here $\mathcal{P},\mathcal{N}$ denote quarks. Note that all neutrino-induced amplitudes are simply given by G_V or G_A.

Using the coupling parameters $C_V^{\mathcal{P}}$ and $C_V^{\mathcal{N}}$, we may now obtain the effective charge Q_W characterizing atomic parity violation.[31] This is given by

$$Q_W=-2[(2C_V^{\mathcal{P}}+C_V^{\mathcal{N}})Z+(2C_V^{\mathcal{N}}+C_V^{\mathcal{P}})(A-Z)] \tag{40a}$$
$$=\frac{(K_1-K_2)(\sin2\beta)[2Z-(3\gamma+1)A]}{(3\gamma+1)^{1/2}} \tag{40a}$$
$$=\frac{2Z-(3\gamma+1)A}{(3\gamma+1)}\left[\frac{(\eta-\zeta)}{\zeta+\eta-\zeta^2}\right] \tag{40b}$$
$$=Q_W(\mathrm{SU}(2)\times\mathrm{U}(1))_\gamma\left(\frac{\eta-\zeta}{\zeta+\eta-\zeta^2}\right), \tag{40c}$$

where

$$Q_W(\mathrm{SU}(2)\times\mathrm{U}(1))_\gamma=[2Z-(3\gamma+1)A]/(3\gamma+1)$$

is the value of Q_W obtained for the $\mathrm{SU}(2)\times\mathrm{U}(1)$ theory (Ref. 4) with $\sin^2\theta_W=3\gamma/(6\gamma+2)$ See Sec. IV).

IV. THE TWO SPECIAL AND THE TWO AMUSING CASES

It is first important to notice the properties of two special cases:

Case A: *The case of vanishing or small fermion masses and small W_L-W_R mixing* $a^2/(c^2+a^2)\to0$, or $\zeta=c^2/(c^2+a^2)\to1$. It turns out that one obtains good agreement with the entire body of present experimental values of neutral-current parameters for the case $\zeta\approx1$ (i.e., $a^2\ll c^2$). In order to understand the results for this case, which we discuss in detail later in this section, it is first helpful to note one (remarkable) general property of the special limit ($\zeta=1$, i.e., $a^2=0$), in which W_L-W_R mixing as well as fermion masses vanish.[32] [Appropriately, this limit will be referred to as the global (γ_5-invariant) chiral-symmetry limit.]

Theorem I: *The chiral-symmetry theorem*. In the limit of exact global chiral symmetry ($a^2\to0$ or $\zeta\to1$), even though the masses of the two neutral eigenstates N_1 and N_2, the mixing angle β, and the atomic parity-violation parameter Q_W vary sensitively with b^2 [i.e., with $\eta=b^2/(c^2+a^2)$], *the amplitudes of all neutrino-induced processes initiated by left-handed neutrinos (or right-handed antineutrinos) are totally independent of the vacuum expectation values of "right-handed" scalar fields such as B*. In the present case they are independent of b^2, depending just on the single parameter $\gamma={f_{15}}^2/g^2$. [By "right-handed scalar field," we simply mean a field which couples to W_R's, but not to W_L's, as for example $B=(1,2+2,\overline{4})$.]

The exact value of a^2 cannot, of course, be zero, since fermions do have finite mass. However,

TABLE I. Predictions of all neutral-current parameters for the nonchiral case ($a^2 \neq 0$, i.e., $\zeta \neq 1$), $r = 0.44$.

Parameters	$\zeta = 0.9$					$\zeta = 0.8$				
$(1-\zeta)/\eta \equiv y$	0.5	0.16	0.10	0.05	0.03	0.5	0.16	0.10	0.05	0.03
R_ν	0.22	0.25	0.26	0.26	0.27	0.22	0.24	0.25	0.26	0.27
$R_{\bar\nu}$	0.24	0.31	0.34	0.36	0.37	0.26	0.30	0.31	0.34	0.35
R	0.37	0.42	0.44	0.46	0.46	0.40	0.38	0.41	0.43	0.45
$C_{\bar\nu_\mu e}$	0.09	0.13	0.14	0.15	0.16	0.07	0.12	0.13	0.14	0.15
$C_{\nu_\mu e}$	0.05	0.09	0.10	0.11	0.12	0.03	0.07	0.09	0.10	0.11
$C_{\bar\nu_e e}/0.57$	1.22	1.17	1.16	1.16	1.15	1.29	1.20	0.18	1.17	1.16
R_ν^{e1}	0.08	0.10	0.11	0.11	0.12	0.07	0.09	0.10	0.11	0.11
$R_{\bar\nu}^{e1}$	0.08	0.14	0.15	0.17	0.18	0.07	0.11	0.13	0.15	0.17
R^{e1}	0.35	0.45	0.48	0.50	0.51	0.30	0.39	0.43	0.47	0.50
$(m_{W_L^+}/m_{N_1})^2$ (GeV^2)	2.39	1.12	0.45	0.77	0.74	1.80	0.94	0.79	0.74	0.73
$(m_{W_L^+}/m_{N_2})^2$ (GeV^2)	0.62	0.55	4.07	0.29	0.19	0.66	0.60	0.47	0.30	0.19
Q_W	360	0.60	−13.0	−65	−94.0	230	36	−24	−70	−96

as long as $a^2 \ll c^2$ (i.e., $\zeta \approx 1$), the amplitudes of neutrino-induced processes continue to remain insensitive to variations of b^2 (or η), varying just with r, on account of the result stated above. This is in fact borne out by our calculations (see Table I).

While the validity of the result stated above may be deduced on general grounds, it may be verified simply by noting the values acquired by the relevant parameters G_V, G_A, K_1, K_2, and Q_W at $\zeta = 1$. These may be obtained from Eqs. (30), (37), (38), and (40) and are listed below.

$$G_V(\zeta = 1) = 1/(3r+1), \tag{41}$$

$$G_A(\zeta = 1) = 1, \tag{42}$$

$$K_i(\zeta = 1) = [m_{W_L^+}{}^2/m_{N_i}{}^2]_{(\zeta=1)} = 4\{(3r+2)(1+\eta) \mp [9r^2(1+\eta)^2 + 4(3r+1)(1-\eta)^2]^{1/2}\}^{-1}, \tag{43}$$

$$Q_W(\zeta = 1) = \frac{[2Z - (3r+1)A]}{3r+1}\left(\frac{\eta - 1}{\eta}\right). \tag{44}$$

[In Eq. (43), the upper (lower) sign refers to the case $i = 1$ ($i = 2$).] Note that G_V and G_A and therefore the amplitudes of all neutrino-induced processes [see Eq. (39)] are independent of b^2 (or η) at $\zeta = 1$; they depend on the single parameter $r = (f_{15}{}^2/g^2)$. On the other hand, the masses of N_1 and N_2 (given by K_1 and K_2) as well as the atomic parity-violation parameter[33] Q_W depend on r as well as b^2 (i.e., η) even at $\zeta = 1$ (the chiral limit $a^2 = 0$).

Case B: The case of $SU(2) \times U(1)$ limit [$b^2 \gg (c^2 + a^2)$, or $\eta \gg 1$]. We present the results for the case $\eta \gg 1$ by assuming $r \neq 0$ (i.e., $\eta r \gg 1$). *As we will show, the results in this case are independent of ζ (varying over its full range from 0 to 1).* Except for the mass of N_2, which becomes superheavy depending upon the value of η, all relevant neutral-current parameters are independent of η as well. They depend once again (as in case A) on the single parameter $r = f_{15}{}^2/g^2$. Using Eqs. (14)–(19), we obtain in the limit $\eta \gg 1$ and $\eta r \gg 1$,

$$m_{N_1}{}^2 = m_{W_L^+}{}^2\left(\frac{6r+2}{3r+2}\right)\left[1 + O\left(\frac{1}{\eta}\right)\right], \tag{45}$$

$$m_{N_2}{}^2 = \tfrac{1}{2}m_{W_L^+}{}^2(3r+2)\eta\left[1 + O\left(\frac{1}{\eta}\right)\right] \gg m_{W_L^+}{}^2, \tag{46}$$

$$\tan 2\beta = +\tfrac{2}{3}(3r+1)^{1/2}/r, \tag{47}$$

$$A = W_L^3 \sin\theta_W + \cos\theta_W B^0, \tag{48}$$

$$N_1 \to W_L^3 \cos\theta_W - \sin\theta_W B^0, \tag{49}$$

$$N_2 \to \frac{(\frac{3}{2}r)^{1/2}V_{15} + W_R^3}{(\frac{3}{2}r+1)^{1/2}}, \tag{50}$$

where

$$B^0 = \frac{f_{15}W_{3R} - (\frac{2}{3})^{1/2}gV_{15}}{(f_{15}{}^2 + \frac{2}{3}g^2)^{1/2}}, \tag{51}$$

$$\sin^2\theta_W = \frac{3r}{6r+2}. \tag{52}$$

Note that the expression (48) for the photon field is the same as we had before [see Eq. (14a)]. Thus we see that the theory reduces in the limit $\eta \gg 1$ to the $SU(2) \times U(1)$ theory with the correspondence: B_μ^0 is the U(1)-gauge field, N_1 is the Z^0 with a mass comparable to that of $W_L^\pm$, while N_2 is superheavy. The familiar angle $\sin^2\theta_W$ is given by Eq. (52) with

$$m_Z{}^2 = m_{N_1}{}^2 \approx m_{W_L^+}{}^2\left(\frac{6r+2}{3r+2}\right) = m_{W_L^+}{}^2/\cos^2\theta_W.$$

Note in the bare limit of a unified theory, which incorporates $\mathcal{G} = SU(2)_L \times SU(2)_R \times SU(4)'_{L+R}$, one would obtain[34] $(\sin^2\theta_W)_{\text{bare}} = 3r/(6r+2) = \frac{3}{8}$, since $r = f_{15}{}^2/g^2 = 1$ in this bare limit.

Finally, we see that the relevant parameters G_V, G_A, and Q_W (in the limit $\eta \gg 1$) are given by

$$G_V = \frac{\eta + 2(1-\zeta)}{(3r+1)(\zeta+\eta-\zeta^2)} \xrightarrow[(\zeta=0 \text{ to } 1)]{\eta \gg 1} \frac{1}{3r+1}, \tag{53}$$

$$G_A = \frac{\eta}{\zeta+\eta-\zeta^2} \xrightarrow[(\zeta=0 \text{ to } 1)]{\eta \gg 1} 1, \tag{54}$$

$$Q_W = \frac{2Z-(3r+1)A}{3r+1}\left(\frac{\eta-\zeta}{\zeta+\eta-\zeta^2}\right) \xrightarrow[(\zeta=0 \text{ to } 1)]{\eta \gg 1} \frac{2Z-(3r+1)A}{3r+1}. \tag{55}$$

Thus, the most amusing feature is that the expressions for G_V and G_A obtained in the SU(2) × U(1) limit are identical to those obtained in case A (the chiral-symmetry limit: $a^2 \to 0$ or $\zeta = 1$); the two cases differ, in general (depending upon the value of η), only in their predictions of the masses of N_1 and N_2 as well as of the atomic parity-violation parameter Q_W. *They agree entirely in all their predictions as regards amplitudes for neutrino-induced processes.*

Case C: An amusing special case $[r = f_{15}^2/g^2 = \frac{1}{3}]$. The two cases A and B discussed above represent all the principal features of the left-right-symmetric gauge structure insofar as its predictions for neutral-current phenomena are concerned. We note in passing a third amusing special case which may come under either case A or case B, or both[35] in view of the remarkably simple forms which the basis vectors (A_μ, Y_1, Y_2), and consequently the eigenvectors $(A_\mu, N_1, \text{and } N_2)$, acquire in this case. This case corresponds to the coupling-constant ratio[36]:

$$r = f_{15}^2/g^2 = \tfrac{1}{3}. \tag{56}$$

With $r = \frac{1}{3}$, the mixing angle α determining the composition of the photon and Y_2 [using Eqs. (13a) and (14)–(18)] is given by

$$\sin^2\alpha = \tfrac{2}{3}e^2/f_{15}^2 = \tfrac{1}{2}. \tag{57}$$

Thus the angle α is simply 45°. Consequently,

$$A = \frac{1}{\sqrt{2}}\left(\frac{W_L^3 + W_R^3}{\sqrt{2}}\right) - \frac{1}{\sqrt{2}}V_{15},$$

$$Y_1 = \frac{1}{\sqrt{2}}(W_L^3 - W_R^3), \tag{58}$$

$$Y_2 = \frac{1}{\sqrt{2}}\left(\frac{W_L^3 + W_R^3}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}}V_{15}.$$

The Y_1-Y_2 mass matrix reduces to

$$M_{(Y_1,Y_2)}^2 = \tfrac{1}{4}g^2 \begin{bmatrix} Y_1 & Y_2 \\ \tfrac{1}{2}(4a^2+b^2+c^2) & -\frac{1}{\sqrt{2}}(b^2-c^2) \\ -\frac{1}{\sqrt{2}}(b^2-c^2) & (b^2+c^2) \end{bmatrix}. \tag{59}$$

TABLE II. Prediction of neutrino parameters in the chiral limit ($a^2 = 0$, or $\zeta = 1$) as a function of the coupling-constant ratio $r = f_{15}^2/g^2$. Chiral-limit predictions of *neutrino parameters* of the left-right-symmetric theory are *identical* to those of the SU(2)$_L$ × U(1) theory with the correspondence $\sin^2\theta_W = 3r/(6r+2)$ (see text). They are independent of $\eta = b^2/(c^2+a^2)$. The semileptonic experimental parameters are defined by $R_\nu = (\sigma_{NC}^{\nu N}/\sigma_{CC}^{\nu N})$, $R_{\bar\nu} = (\sigma_{NC}^{\bar\nu N}/\sigma_{CC}^{\bar\nu N})$, $R = (\sigma_{NC}^{\bar\nu N}/\sigma_{NC}^{\nu N})$, and similarly R_ν^{e1}, $R_{\bar\nu}^{e1}$, and R^{e1}, where NC and CC refer to neutral- and charged-current cross sections, respectively. The parameters $C_{\nu_\mu e}$, $C_{\bar\nu_\mu e}$, $C_{\bar\nu_e e}$, and $C_{\bar\nu_e e}$ have been defined in Ref. 37. The predictions of elastic-scattering parameters are for a choice of $M_A^2 = 1.15$ GeV^2, where M_A^2 characterizes the axial vector form factor. The predictions would be somewhat lower (by about 10–15%) if we choose $M_A = 1$ GeV.

	Predictions						Experiment	
Parameters	$r=0.11$	$r=0.22$	$r=0.33$	$r=0.44$	$r=0.77$	$r=1$	GGM (Ref. 40)	CITF (Ref. 41)	HPWF (Ref. 42)
R_ν	0.39	0.33	0.30	0.27	0.24	0.23	0.25 ± 0.04	0.24 ± 0.02	0.29 ± 0.04
$R_{\bar\nu}$	0.41	0.39	0.39	0.40	0.42	0.44	0.39 ± 0.06	0.34 ± 0.09	0.39 ± 0.10
R	0.35	0.39	0.44	0.48	0.58	0.64	0.59 ± 0.14	...	$\leq 0.61 \pm 0.25$
$C_{\bar\nu_\mu e}$	0.11	0.12	0.14	0.17	0.22	0.25	$0.11^{+0.21}_{-0.08}$ GGM (Ref. 40)	0.54 ± 0.17 Aachen Padova (Ref. 43)	
$C_{\nu_\mu e}$	0.25	0.18	0.14	0.13	0.11	0.11	0.24 ± 0.12	Aachen Padova (Ref. 43)	
$C_{\bar\nu_e e}/0.57$	0.58	0.81	1.00	1.15	1.45	1.58	0.87 ± 0.25 (1.5 ~ 3 MeV)		
							1.70 ± 0.44 (3 ~ 4 MeV)	Reines *et al.* (Ref. 44)	
								CIR (Ref. 45)	HPWF (Ref. 46)
R_ν^{e1}	0.19	0.15	0.13	0.12	0.10	0.10	...	0.23 ± 0.09	0.17 ± 0.05
$R_{\bar\nu}^{e1}$	0.19	0.17	0.18	0.19	0.22	0.24	...	...	0.20 ± 0.10
R^{e1}	0.34	0.39	0.46	0.53	0.72	0.81	...	...	0.4 ± 0.02

The point $r=\frac{1}{3}$ is a point of good agreement between theory and experiments (as may be seen from Table II). (Note, if in addition to $r=\frac{1}{3}$, one furthermore imposes $a^2=(b^2+c^2)/4$, the diagonal elements of the (Y_1, Y_2) mass matrix become equal, so that the eigenstates N_1 and N_2 are given by the simple combinations $(Y_1 \pm Y_2)/\sqrt{2}$. A second interesting feature of this choice is that the vector coupling (M_V^e) for the process $\nu_\mu e \to \nu_\mu e$ vanishes [see Eq. (39)]. This implies in turn that the two processes $\nu_\mu e \to \nu_\mu e$ and $\bar{\nu}_\mu e \to \bar{\nu}_\mu e$, both proceeding only through axial-vector electron current, have equal cross sections for the case $r=\frac{1}{3}$.

Case D: The case of exact neutral-current parity conservation $(b=c)$. If $b=c$, as noted in Secs. II and III, Y_1 and Y_2 are the eigenstates of the "neutral" mass matrix. In this case neutral-current interactions (of electrons, quarks, as well as of neutrinos) are exactly parity conserving, since Y_1 and Y_2 are coupled to axial-vector and vector currents, respectively. This is the case considered by MS (Ref. 16). In this case, as noted before, ν-$\bar{\nu}$ distinction arises not becasue of the intrinsic parity-violating nature of the neutral-current interactions, but because available neutrinos are only left-handed (and antineutrinos are only right-handed). Note this particular case can overlap with case A (the chiral case), but it is exclusive of case B [the SU(2) × U(1) limit], for which $b^2 \gg (c^2+a^2)$. This particular case $(b=c)$ of exact atomic parity conservation, it turns out, does not permit the possibility of a rather light neutral gauge particle $(m_{N_1} < m_{W_L^+})$ (see discussions in Sec. V).

V. PHENOMENOLOGY: DISCUSSION OF RESULTS

With the amplitudes and K_1, K_2, and Q_W listed in Sec. III, it is straightforward to calculate the various neutral-current parameters[37] (R_ν, $R_{\bar{\nu}}$, R, $C_{\bar{\nu}_\mu e}$, $C_{\nu_\mu e}$, $C_{\bar{\nu}_e e}$, R_ν^{e1}, $R_{\bar{\nu}}^{e1}$, and R^{e1}) using the quark-parton model for the semileptonic processes. For simplicity, we neglect the contributions of the anti-quark-partons[38] in the sea. The results exhibited in Tables I, II, and III as well as in Figs. 1–8 represent the principal features[39] of our analysis. We draw attention to the following features of our results:

(1) We see first of all from Tables I and II that the point of *exact or approximate chiral symmetry* [i.e., $1-\zeta \equiv a^2/(c^2+a^2) \approx 0$], together with the coupling-constant ratio r being in the range ≈0.33 to 0.66 (see below), is a point of good agreement between theory and experiments[40–46] involving *all* present neutral-current processes. One obtains in this case a fit[47] simultaneously for the neutrino data as well as for the electron atomic parity experiments, which presently seem to indicate[1,2] a small value for Q_W.

The question arises, how far from the chiral limit ($\zeta=1$ corresponding to $a^2=0$) may one depart and yet be in accord with the experiments. The answer we find is interesting: one may choose $\zeta \equiv c^2/(c^2+a^2)$ sensibly lower than unity (i.e., $a^2 \sim c^2$) if one wishes and yet be in accord with all the neutrino data provided, however, that the special combination[30]

$$y \equiv (1-\zeta)/\eta = a^2/b^2 \tag{60}$$

is sufficiently small (<0.15). This may be inferred partly from Table I and Figs. 2, 3, and 4, which show that the values of $R_{\bar{\nu}}$ and in particular R become lower[48] than those permitted by the present data, if y exceeds nearly 0.15 with ζ = either 0.9 or 0.8. [It may be verified that the neutral-current cross-section *ratio* R is only a function of the coupling-constant ratio r and the parameter y, but it is *independent* of $x \equiv (\eta-\zeta)/\zeta = (b^2-c^2)/c^2$ and thus is an especially useful quantity for

TABLE III. Predictions of the masses of N_1 and N_2 and the atomic parity-violation parameter in the chiral limit ($a^2=0$ or $\zeta=1$) as a function of $(\eta-\zeta)/\zeta$.

$r=\frac{f_{1\mathfrak{g}}^2}{g^2}$	$\frac{\eta-\zeta}{\zeta}=\frac{b^2-c^2}{c^2}$	$K_1=\frac{M_{W_L^+}^2}{M_{N_1}^2}$	$K_2=\frac{M_{W_L^+}^2}{M_{N_2}^2}$	Q_W	$M_{W_L^+} \simeq 53(1+1/3r)^{1/2}$
0.33	−0.5	1.64	0.61	126	75 GeV
	−0.375	1.36	0.59	75	
	0.0	1.00	0.50	0.0	
	1.0	0.82	0.34	−63	
SU(2) × U(1) limit	99	0.75	0.007	−125	$\sin^2\theta_W=0.25$
0.44	−0.5	1.61	0.53	138	70 GeV
	−0.375	1.34	0.51	83	
	0.0	1.00	0.43	0.0	
	1.0	0.81	0.27	−69	
SU(2) × U(1) limit	99	0.72	0.006	−136.5	$\sin^2\theta_W=0.29$

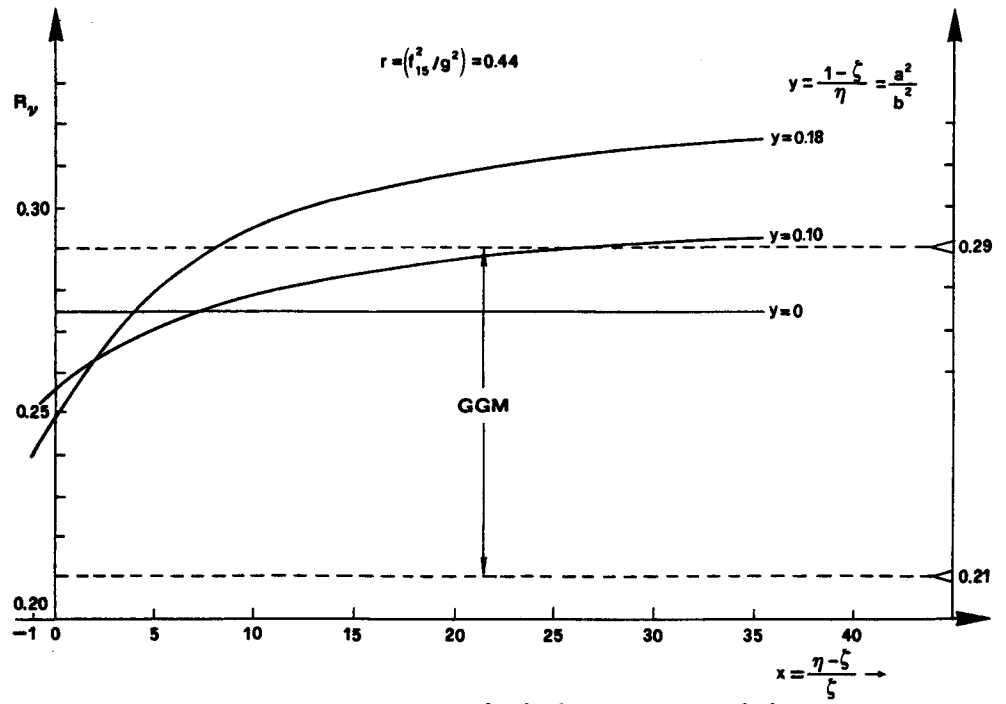


FIG. 2. Variation of R_ν as a function of $x=(\eta-\xi)/\xi=(b^2-c^2)/c^2$ for $y=(1-\xi)/\eta=a^2/b^2=0.0$, 0.1, and 0.18 with the coupling-constant ratio r fixed at 0.44.

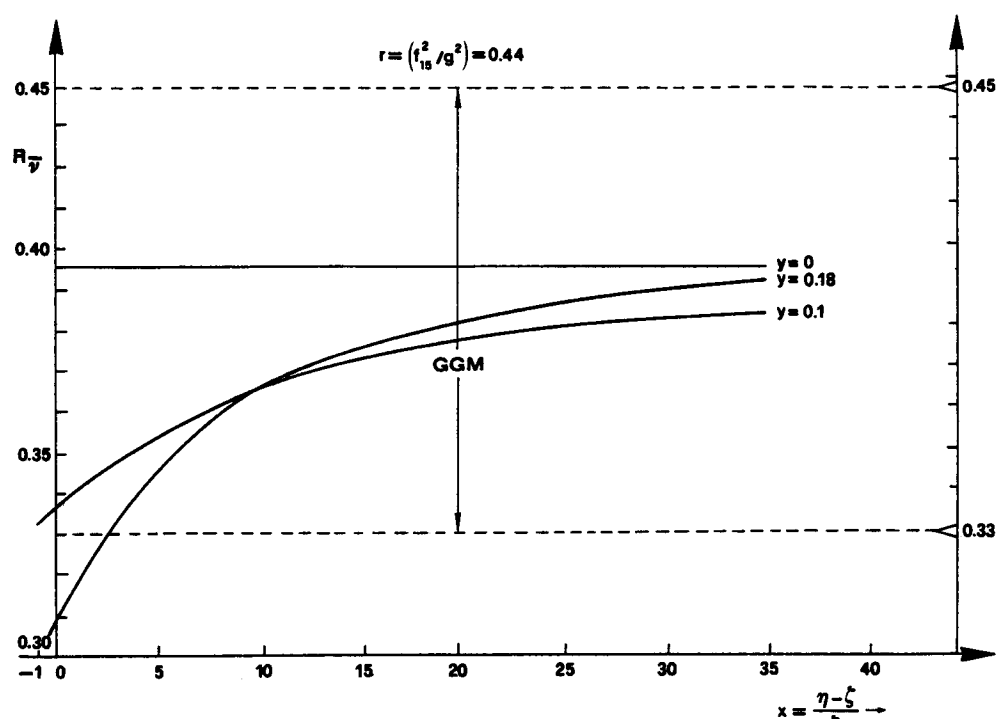


FIG. 3. Variation of $R_{\bar{\nu}}$ as a function of x for $y=0.0$, 0.1, and 0.18 with r fixed at 0.44.

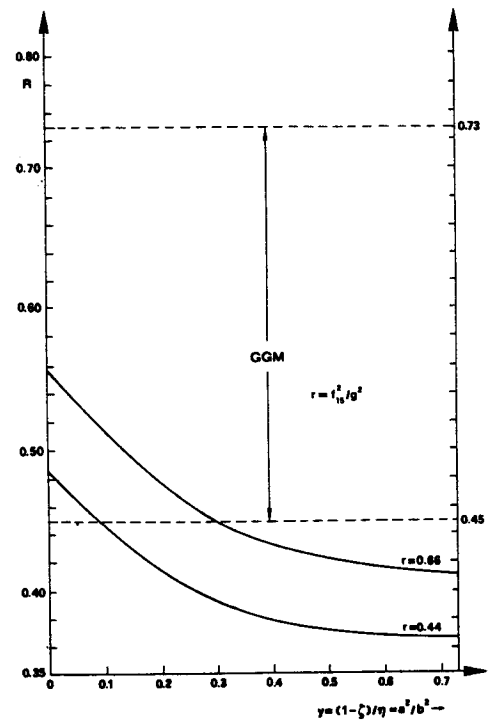


FIG. 4. Variation of $R=(\sigma^{\bar{\nu}N}_{NC}/\sigma^{\nu N}_{NC})$ as a function of y with $r=0.44$ and 0.66. Note that R is *independent* of $x=(\eta-\zeta)/\zeta=(b^2-c^2)/c^2$.

comparison of theory with experiment.] The condition that y be small ($\ll\frac{1}{6}$) can no doubt be satisfied even with arbitrarily small $\zeta\to 0$ by choosing η to be sufficiently large, i.e.,

$$\eta>6(1-\zeta). \tag{61}$$

For example, if ζ is as small as 0.5, one would need $\eta>3$; this is simply to explain the neutrino data. The price one must pay in this case, however (if it is a price[2]), is that the atomic parity-violation parameter Q_W would rapidly approach its $SU(2)_L\times U(1)$ value for $\eta\gg 1$ [see Eq. (40c) and our discussions in Sec. IV] in apparent contradiction with the present data; simultaneously the mass of N_2 would become progressively heavier ($\gg m_{W_L^+}$) with N_1 remaining relatively light but still *heavier* than W_L^+.

The precise constraint which may be obtained on the chiral-symmetry-breaking parameter $(1-\zeta)$ by the above considerations can only be laid down after one knows the value of Q_W from experiments together with a reliable atomic-physics calculation[2] for the relevant atom. However, to demonstrate the point, let us assume *for illustration only* that the true value of Q_W is at least a factor of 3 lower *in magnitude* than its $SU(2)\times U(1)$ value. This would imply, using Eq. (40c), that the ratio $Q_W/Q_W(SU(2)\times U(1))=(\eta-\zeta)/(\zeta+\eta-\zeta^2)$ lies between $-\frac{1}{3}$ and $+\frac{1}{3}$, which in turn would imply (given *either* sign for Q_W) that $\eta<(2\zeta-\frac{1}{2}\zeta^2)$. Combining this with the constraint that the neutrino data seem to require $6(1-\zeta)<\eta$, as stated above, we would deduce

$$(8\zeta-\zeta^2/2)>6$$

or

$$1-\zeta=\frac{a^2}{c^2+a^2}<0.2\,.$$

We stress once again that this constraint should be taken at present as an illustration only, since the true value of Q_W is yet an unknown quantity.

We draw the following important conclusion from the above discussion: If Q_W indeed turns out to be rather small compared to its $SU(2)\times U(1)$ value, and if the neutrino parameters retain their present values, one can well account for both these features within the left-right–symmetric

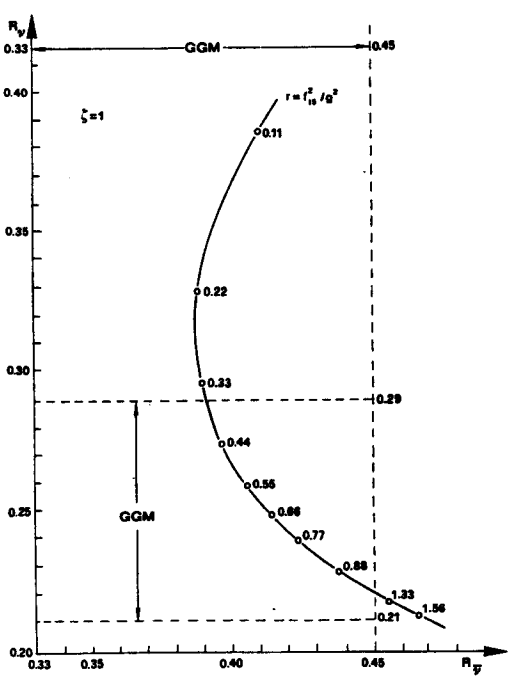


FIG. 5. R_ν versus $R_{\bar{\nu}}$ for varying coupling-constant ratio r at the chiral-symmetric point $\zeta=c^2/(c^2=a^2)=1$, where all neutrino amplitudes are functions of the single parameter r with $\sin^2\theta_W=(3r)/(6r+2)$. The allowed range of the coupling-constant ratio r (following from comparison of the predicted values with experiments) is not altered appreciably by departing from the chiral limit ($\zeta=1$) and varying $\eta=b^2/(c^2+a^2)$ as well.

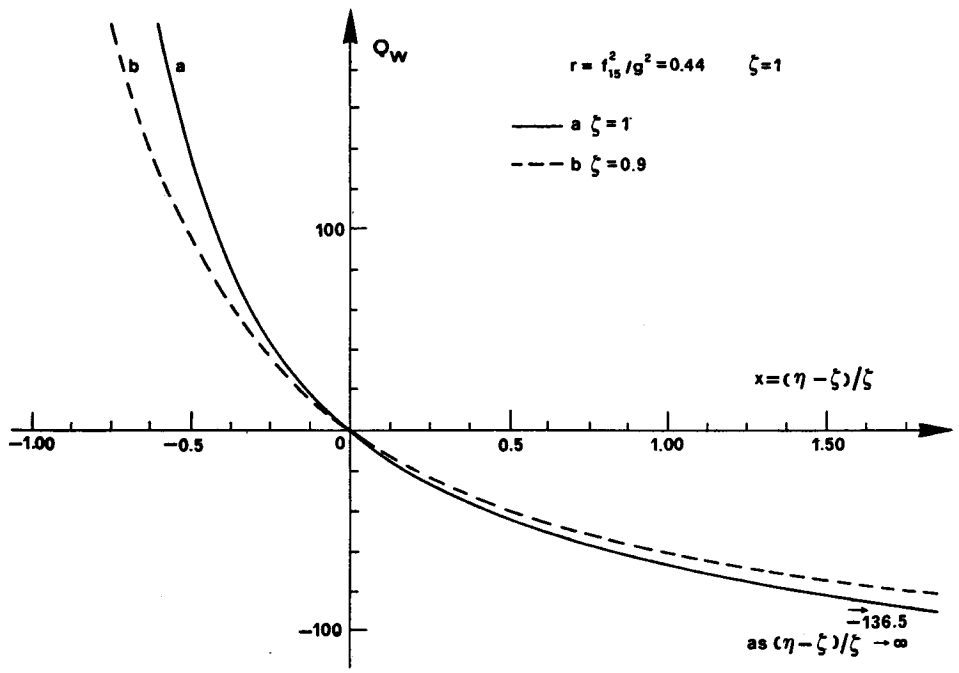


FIG. 6. Variation of the atomic parity-violation parameter Q_W with $x = (\eta - \zeta)/\zeta = (b^2 - c^2)/c^2$ for the cases of exact or approximate chiral symmetry (i.e., $1 - \zeta = 0$ and 0.1) with fixed coupling-constant ratio r.

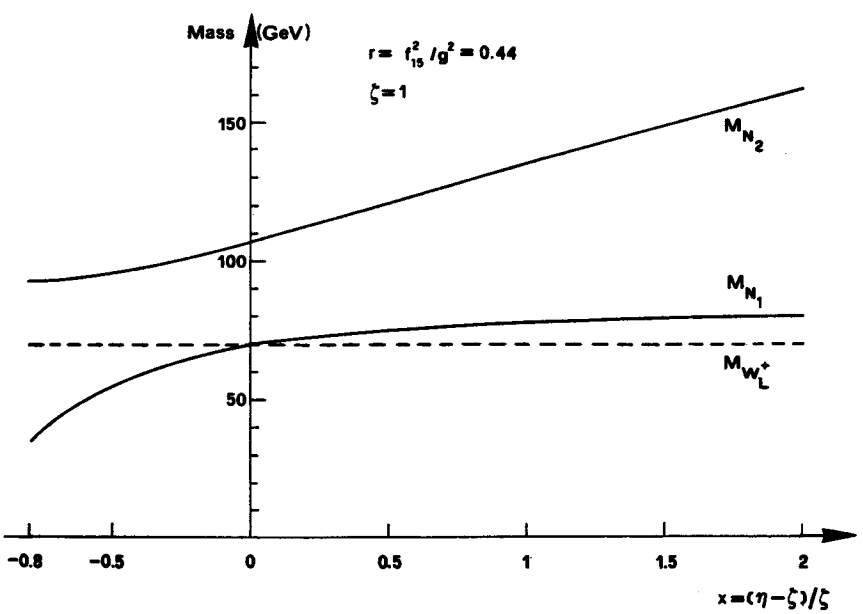


FIG. 7. Variation of the predicted masses of the two neutral-gauge particles N_1 and N_2 with $x = (\eta - \zeta)/\zeta$ for the chiral-symmetric point ($\zeta = 1$) with fixed coupling-constant ratio r. These predictions are insensitive to small departures from the chiral-symmetric point (i.e., ζ varying, say, between 0.8 to 1) *even* at relatively small x.

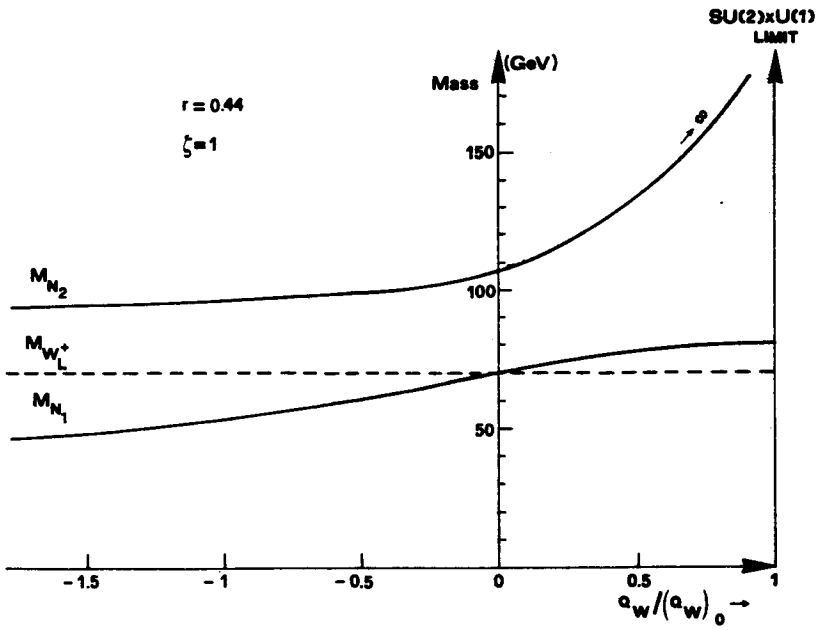


FIG. 8. Predicted masses of (N_1 and N_2) as a function of the ratio of atomic parity-violation parameters $Q_W/(Q_W)_0$. Q_W and $(Q_W)_0$ are the values of the atomic parity-violation parameters for Bi for the left-right-symmetric model (Ref. 6) and the $SU(2)_L \times U(1)$ model (Ref. 4), respectively. Expressions for these quantities are given in the text [see Eqs. (40) and (55)].

theory; however, *in this case the solution of the theory must lie near the chiral-symmetric limit* ($a^2 \ll c^2$), where the limit $a^2 = 0$ corresponds to the vanishing W_L-W_R mixing as well as the vanishing fermion masses.

(2) We noted above that agreement between our predictions and the neutrino data is obtained provided the coupling-constant ratio $r = f_{15}^2/g^2$ lies in a limited range. Examining Table II and Fig. 4, one may deduce (this holds for $\zeta < 1$ as well) *a guiding* range:

$$0.33 \lesssim r = f_{15}^2/g^2 < 0.66 . \tag{62}$$

If the coupling-constant ratio r is lower than 0.30, the neutrino parameter R_ν becomes too high, exceeding 0.30, which lies above the upper limits of Gargamelle (GGM) and Caltech-Fermilab (CITF)-values. If, on the other hand, the coupling-constant ratio r exceeds 0.66, $R_{e1} = \sigma_{el}^{\bar{\nu}p}/\sigma_{el}^{\nu p}$ becomes too high, exceeding the upper limit ≈ 0.6 of the present Harvard-Pennsylvania-Wisconsin-Fermilab (HPWF) result. [This last deduction should, however, be taken with caution, since the theoretical prediction of elastic-scattering parameters depend[47] upon $(\nu, \bar{\nu})$ spectra and the mass characterizing the axial-vector form factor.]

It is remarkable that the range of the coupling-constant ratio, exhibited above, deduced phenomenologically, lies well within the rather "limited" range of the same, deduced on *theoretical grounds*[24] based on renormalization-group considerations, which yield

$$\tfrac{1}{3} e^2 \lesssim f_{15}^2 \lesssim \tfrac{5}{3} e^2 . \tag{63}$$

Together with Eq. (13a), this implies

$$\tfrac{1}{3} \lesssim r = f_{15}^2/g^2 \lesssim 1 . \tag{64}$$

The lower limit corresponds to $f_s/g \to \infty$ and the upper to the bare value ($f_s = f_{15} = g$).

(3) The variation of Q_W as a function of[30] $x \equiv (\eta - \zeta)/\zeta = (b^2 - c^2)/c^2$ is exhibited in Fig. 6. As we see, it passes through zero at $x = 0$, approaching its $SU(2) \times U(1)$ value in the limit $x \to \infty$ with $\zeta \sim 1$ (i.e., $b^2 \gg c^2$, with $c^2 \neq 0$), while for negative x ($b^2 < c^2$), it rises rapidly acquiring unbounded positive values. We find that large negative x (i.e., $x < -0.6$, or $b^2 < 0.4c^2$) is probably excluded by the present atomic parity experiments, since such x implies $Q_W > +300$. On the other hand, small negative values of x ($-0.5 \lesssim x \equiv (\eta - \zeta)/\zeta \leq 0$ corresponding to $\frac{1}{2} \leq b^2/c^2 \lesssim 1$) are fully allowed by the data on Q_W (as well as by the neutrino data).

(4) If $x > 0$ (i.e., if $b^2 > c^2$), then both N_1 and N_2 remain heavier than W_L^+ (see Fig. 7), with the mass of N_1 coinciding with that of W_L^+ for $b^2 = c^2$ (at $\zeta = 1$, i.e., $a^2 = 0$). On the other hand, *if* $x < 0$

(i.e., *if* $b^2 < c^2$), N_1 *is inevitably lighter than* W_L^+. For example, if the true value of the atomic parity-violation parameter Q_W is positive and as large as 100 (this is consistent with the present data, Refs. 1 and 2), then from Table III or Fig. 6 we may deduce[49] that the parameter $x = (\eta - \zeta)/\zeta$ should be ≈ -0.44, and correspondingly (from Fig. 7) with $r = 0.44$, we would obtain[50]

$$m_{N_1} \approx 56 \text{ GeV}, \quad m_{W_L} + \approx 70 \text{ GeV}. \tag{65}$$

By contrast, for the SU(2) × U(1) theory with $r = 0.44$ [i.e., $\sin^2\theta_W = 3r/(6r+2) \approx 0.29$], we obtain

$$m_{Z^0} \approx 83 \text{ GeV}, \quad m_{W_L} + \approx 70 \text{ GeV}. \tag{66}$$

Theorem II. The left-right-symmetric theory offers the distinct advantage that one of the neutral gauge particles is necessarily lighter than the Z^0 of the SU(2)×U(1) theory. If $b^2 < c^2$, it will be lighter than the charged W_L^+ as well.

(5) In our analysis, we have neglected for simplicity the effect of W_L-W_R mixing on the composition and masses of the charged eigenstates. [Compare the mass matrix (9b) with Eq. (26).] This is justifiable *a posteriori* within the four-flavor model[6]: First, because[51] with only four flavors one must have $m_{W_R^+}^2 \geq 10 m_{W_L^+}^2$ [following Eq. (11) and the remark in Ref. 26] to account for the observed predominance of $V-A$ over $V+A$ [$m_{W_L^+}^2$ and $m_{W_R^+}^2$ here refer to the diagonal elements of (9b)]; second, because, as noted above, we found *a posteriori* that the solution with approximate chiral symmetry ($a^2 \ll c^2$) may even be favored by the data.

Extensions of the four-flavor model[6] to models containing more than four flavors and/or more than four colors, while still *retaining* the left-right-symmetric gauge structures[6] $\mathcal{G}$ or $\bar{\mathcal{G}}$ (as subgroups) are straightforward (see, for example, the mirror model Ref. 25 for such an extension). What we wish to stress here is that *such* extensions do not alter any of our considerations regarding neutral-current interactions. They would in general matter for charged-current interactions. We would enter into a more detailed study of this possibility elsewhere.

To conclude, the left-right-symmetric theory[6] accounts for all the present experimental neutral-current parameters including the phenomena of possibly small atomic parity violation—in terms of three parameters in general [$r = f_{15}^2/g^2$, $\eta = b^2/(c^2+a^2)$, and $\zeta = c^2/(c^2+a^2)$]. The allowed range for the coupling-constant ratio r (≈ 0.33 to 0.66) deduced phenomenologically falls within the theoretically deduced[24] rather limited range for the same. In addition to r, the combinations $(1-\zeta)/\eta = a^2/b^2$ and $(\eta - \zeta)/\zeta = (b^2 - c^2)/c^2$ also get constrained severely by the present data in spite of their large errors. Thus the left-right-symmetric theory can be well tested experimentally once the body of independent neutral-current measurements settles experimentally. There are at least eight such measurements involving scattering-type experiments only. Including the masses of the two neutral particles, this number is ten.

The intricacies of the left-right-symmetric theory are remarkable. Depending upon the pattern and magnitudes of the parameters of spontaneous symmetry breaking,[6] it offers two interesting alternative limits: (i) the chiral-symmetry limit ($a^2 \ll c^2$) with $b^2 \sim c^2$ and (ii) the SU(2) × U(1) limit ($b^2 \gg c^2 + a^2$). In either case, the neutrino processes are described in terms of the single parameter r. The predictions of the two cases for these processes in fact coincide [in the extreme limit of $a^2 \to 0$ for case (i)]; they differ from each other only in respect to electron atomic parity experiment and the masses of the two neutral gauge particles N_1 and N_2. The first case predicts the existence of two relatively light neutral gauge particles[6] N_1 and N_2, and subject to $b^2 < c^2$, it even offers the exciting possibility that one of the neutral gauge particles can be considerably lighter than the charged W_L^+. The second case generates the single relatively light familiar Z^0 (heavier than W_L^+). Which of these two limits of the left-right-symmetric theory[6] is chosen by nature, if any, appears to depend crucially upon a knowledge of the atomic parity-violation parameter Q_W.

ACKNOWLEDGMENTS

We are grateful to Vernon Barger and Victor Elias for most helpful discussions. We thank Sunanda Banerjee, Bahman Hafizi, Betty Moynihan, and Kenneth Powell for their generous assistance in computations.

*Work supported in part by the Science Research Council, United Kingdom and by NSF, Grant No. GP 43662X. On leave of absence (1976–77) from the University of Maryland, College Park, Maryland.

[1]P. Baird, M. Brimicombe, G. Roberts, P. Sanders, D. Starey, E. Forston, L. Lewis, E. Lindahl, and D. Soreide, Nature (London) 264, 529 (1976).

[2]As noted in Ref. 1, the uncertainty in the atomic-physics calculations (on account of configuration mixing) for the amount of parity violation to be expected in Bi^{209} (Ref. 1) is at present so large that no firm conclusions can be drawn from the present level of ex-

perimentation [see R. D. Cowan, S. Meshkov, and S. P. Rosen, report, 1977 (unpublished)]. Thus one must wait for more refined theory and experiments on the problem before judging whether departures from the $SU(2)\times U(1)$ theory as regards atomic parity violation are indeed showing themselves at low energies. For earlier calculations on atomic parity-violation parameters, see M. A. Bouchiat and C. C. Bouchiat, Phys. Lett. 48B, 111 (1974); J. Bernabeu, T. E. O. Ericson, and C. Jarlskog, Phys. Lett. 20, 315 (1974); I. B. Kriplovich, Zh. Eksp. Teor. Fiz. Pis'ma Red. 20, 686 (1974) [JETP Lett. 20, 315 (1974)]; E. M. Henley and L. Wilets, Phys. Rev. A 19, 1911 (1976); M. W. S. M. Brimicombe, C. E. Loving, and P. G. H. Sandars, J. Phys. B 9, L237 (1976).

[3]For measurements of $\sigma(\nu d \to \nu x)$ versus $\sigma(\bar{\nu} d \to \bar{\nu} x)$ see A. Benvenutti *et al.*, Phys. Rev. Lett. 37, 1039 (1976); J. Blietschau *et al.*, Nucl. Phys. B118, 218 (1977); B. C. Barish, Caltech. Report No. CALT-68-544 (unpublished). For measurements of $\sigma(\bar{\nu} p \to \bar{\nu} p)$ versus $\sigma(\nu p \to \nu p)$ see D. Cline *et al.*, Phys. Rev. Lett. 37, 252 (1976); 37, 648 (1976); W. Lee *et al.*, Phys. Rev. Lett. 37, 186 (1976).

[4]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); Abdus Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, edited by N. Svartholm (Almqvist and Wiksell, Stockholm, 1968), p 367, and references therein to the work with J. C. Ward and S. L. Glashow; S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D 2, 1285 (1970).

[5]See, for example, R. M. Barnett, in *Particles and Fields '76*, Proceedings of the Annual Meeting of the Division of Particles and Fields of the American Physical Society, Brookhaven National Laboratory, edited by H. Gordon and R. F. Peierls (BNL, Upton, New York, 1977), p. D 77, for a list of references on standard vectorlike theories based on $SU(2)\times U(1)$ group. It should be noted that it is possible to build a notable class of vectorlike $SU(2)\times U(1)$ theories, as for example the one extracted from E_7 [F. Gürsey and P. Sikivie, Phys. Rev. Lett. 36, 775 (1976); P. Ramond, Nucl. Phys. B110, 214 (1976); P. Sikivie and F. Gürsey, Phys. Rev. D 16, 816 (1977)], in which atomic parity violation can be vanishingly small, and yet there can exist ν-$\bar{\nu}$ distinction in neutral-current scattering experiments (this is unlike the standard vectorlike theories). This class of models, however, appears to suffer from one disadvantage: the strangeness-changing neutral-current processes are not "naturally" suppressed, at least within the $SU(2)\times U(1)$ extractions of these theories. [S. L. Glashow and S. Weinberg, Phys. Rev. D 15, 1958 (1977); see however remarks in P. Sikivie and F. Gürsey, *ibid.* 16, 816 (1977).]

[6]J. C. Pati and Abdus Salam, Phys. Rev. D 10, 275 (1974).

[7]The Abelian factor $U(1)_{L+R}$ within $\mathcal{G} = SU(2)_L \times SU(2)_R \times U(1) \times SU(3)^c_{L+R}$ may be identified with the 15th generator F^c_{15} of $SU(4)^c_{L+R}$, as stressed in Ref. 6 (see footnote 11).

[8]This choice $g_L^{(0)} = g_R^{(0)}$ was proposed in Ref. 6, in order that the origin of parity violation should lie in spontaneous symmetry breaking only. See Ref. 6, p. 276 and footnote 34, where it was stated that, provided $g_L^0 = g_R^0$, the "ratio of the renormalized coupling constants g_L/g_R would differ from unity by calculable finite radiative corrections of order $\alpha \ln(m_{W_R}{}^2/m_{W_L}{}^2)$ to the extent that spontaneous symmetry breaking introduces different masses for W_L and W_R (for example)." The scalar (spontaneously symmetry breaking) multiplets introduced in Ref. 6 (A, B, C) were left-right symmetric, as indeed were the quarks and the leptons. With these choices of multiplets and the gauge structure, the proof that the radiative corrections to (g_L/g_R) can indeed be finite, through an appropriate choice of Higgs parameters, was later carried through in the papers of R. N. Mohapatra and J. C. Pati, Phys. Rev. D 11, 566 (1975); 11, 2558 (1975), and G. Senjanovic and R. N. Mohapatra, Phys. Rev. D 12, 502 (1975).

[9]As long as the left-right discrete symmetry is preserved throughout in the bare gauge Lagrangian *and* in the Higgs potential, except possibly in the scalar *mass terms*, as is the case for the gauge model proposed in Ref. 6, (a) radiative corrections to $(g_L - g_R)/g_R$ remain finite and $O(\alpha)$, and (b) parity nonconservation at low energies of any measurable consequence arise only through the spontaneously induced gauge mass inequality $m_{W_R} \neq m_{W_L}$ [see R. N. Mohapatra and J. C. Pati, Phys. Rev. D 11, 2558 (1975)]. The result on the finiteness of radiative corrections (i.e., $g_L = g_R$ [1 $+ O(\alpha)_{\text{finite}}$]) follows from a theorem due to K. Symanzik [in *Fundamental Interactions at High Energies*, edited by A. Perlmutter *et al.* (Gordon and Breach, New York, 1970)], which states that when a symmetry of the Lagrangian is "softly" broken, vertices with dimensions higher than those of the symmetry-breaking terms suffer only finite renormalizations due to the symmetry-breaking term. Here the symmetry in question is left-right discrete symmetry; the symmetry-breaking terms if any, are scalar-mass terms of dimension two and the vertices have dimensions higher than two.

[10]Parity violation could also arise if the many Cabibbo-like angles, generated by the spontaneously induced Fermi mass matrix within a model with more than four flavors, turn out to be unequal for the left- and right-handed fermions, a possibility which would arise provided one did not insist on manifest left-right symmetry in physical currents. See remarks later.

[11]R. N. Mohapatra and J. C. Pati, Phys. Rev. D 11, 566 (1975).

[12]"Light" in the present context denotes a mass scale of order 50 to 100 GeV.

[13]Barring the insistence that even the physical currents exhibit manifest and natural left-right symmetry, a feature which has guided all the phenomenological constraints such as $m_{W_R^+} \gtrsim 3 m_{W_L^+}$ obtained in Ref. 6 [see recent paper of M. A. B. Bég *et al.*, Phys. Rev. Lett. 38, 1252 (1977), who obtain similar limits and point out that with such insistence left- and right-handed Cabibbo angles θ_L and θ_R (Ref. 6) are indeed equal up to finite radiative corrections strengthening the considerations of Ref. 6].

[14]H. Georgi and S. L. Glashow, Phys. Rev. Lett. 32, 438 (1974). For E_7, see Ref. 5.

[15]It should be understood that there are alternative permissible modes by which the symmetry G can break down to $\mathcal{G}_{\text{obs}}$ other than those exhibited in Fig. 1. For example, G may descend into either $\mathcal{G}_0$ or $\mathcal{G}$ without passing through the intermediate step $\mathcal{G}$. There is

also the distinct possibility that G may pass through $SU(2)_L \times SU(2)_R \times SU(3)'_L \times SU(3)'_R \times U(1)_L \times U(1)_R$ before descending into $\mathcal{G}_{obs}$, where prime denotes color gauging, and $U(1)_{L,R}$ denotes the 15th diagonal generators of $SU(4)'_{L,R}$ (Ref. 7). This latter hierarchy would generate the intriguing alternative of *three* light neutral gauge particles in addition to photon. It will be discussed elsewhere.

[16]R. N. Mohapatra and D. P. Sidhu, Phys. Rev. Lett. 38, 667 (1977). The neutral-current parity-conserving solution of Ref. 6 was noted by H. Fritzsch and P. Minkowski, Nucl. Phys. B103, 61 (1976).

[17]A. De Rújula, H. Georgi, and S. L. Glashow, Harvard report, 1977 (unpublished). A model similar to the six-flavor model considered here has also been considered by Mohapatra and Sidhu (Ref. 16).

[18]This overlap and correspondence is not unexpected, since the origin of the basic considerations of the MS scheme may be traced to our Ref. 6 (see Ref. 13 of the MS paper).

[19]The appearance of two coupling constants g and f, associated with the gauge group $\mathcal{G}$, reflect the passage of a higher unifying symmetry G down to $\mathcal{G}$ through spontaneous symmetry breaking with the effective coupling constants associated with $SU(2)_{L,R}$ and $SU(4)'_{L,R}$ differing from each other through *finite* renormalization effects.

[20]Since $\langle E_L \rangle$ contributes to the masses of the charged $W_L^{\pm}$ only, but not to the neutral W_L^3, allowing for $\langle E_L \rangle \neq 0$ would mean that all our expressions for the neutral-current amplitudes (to be presented in Secs. III and IV) would simply be multiplied by a *uniform* factor κ deviating from unity depending upon $\langle E_L \rangle / m_{W_L^+}$. See remarks later for the experimental implications of $\langle E_L \rangle \neq 0$.

[21]Most notably (2,2,15) or (2+2, 2+2, 15) (see Ref. 6).

[22]We, of course, do not neglect c_1 (for the case of integercharge quarks) in those considerations where its contribution is crucial, such as in the mixing of gluons with the W's as well as in W-X mixing.

[23]J. C. Pati and Abdus Salam, Phys. Rev. D 8, 1240 (1973), J. C. Pati in *Gauge Theory and Modern Field Theory*, edited by R. Arnowitt and P. Nath (M.I.T., Cambridge, 1975), p. 27.

[24]V. Elias, Phys. Rev. D 16, 1586 (1977) and references therein; Q. Shafi, *ibid.* 16, 3293 (1977). Following renormalization-group methods, Elias obtains a $\frac{4}{3} e^2 < f_{15}^2 < 4e^2$, subject to certain assumptions of hierarchical breaking of the unifying symmetry G containing the minimal symmetry $\mathcal{G} = SU(2)_L \times SU(2)_R \times SU(4)'_{L,R}$ of the model of Ref. 6.

[25]For extended models with more than four flavors, for example the mirror model, in which quarks carry eight flavors, see for example J. C. Pati and A. Salam, Phys. Lett. 58B, 333 (1975).

[26]Assuming SU(3)-symmetric matrix elements of vector and axial-vector currents, the present experimental values of g_V and g_A for Λ-β decay seem to suggest that (within the four-flavor model) $(g^2/m_{W_R^+}^2)\sin\theta_R \leq \frac{1}{3}(g^2/m_{W_L^+}^2)\sin\theta_L$, which together with Eq. (11) leads to the inequality $\tan\theta_R \lesssim 3\tan\theta_L$; in turn this implies that right-handed neutrinos may be emitted in strangeness-changing processes ($\Lambda \to p e^- \bar{\nu}_e$ and $K \to \mu + \nu_\mu$, etc.) perhaps as frequently as 10% of the times, even though in strangeness conserving processes their relative rate of emission is known to be $< 10\%$. This may be of experimental consequence in eventually detecting the right-handed neutrinos and left-handed neutrinos and left-handed antineutrinos, which are expected to exist within the model. (See, however, Ref. 13.) In this regard it is worth noting that the model in general may permit the emergence of either (i) light *four-component* ν_e and ν_μ, or (ii) massless two-component $(\nu_{eL}, \nu_{\mu L})$ as well as their two-component right-handed partners $(\nu_{eR}, \nu_{\mu R})$; such masslessness would be preserved provided the eventual mechanism of spontaneous symmetry breaking is constrained to respect γ_5 symmetry for the neutrinos. At any rate it should be stressed that *the doubling of helicities with left-handed fermions being always balanced by the existence of right-handed counterparts, even for neutrinos, is one crucial prediction of the left-right-symmetric theory (Ref. 6).*

[27]Using the mass matrix (9b), (9c), it is easy to verify with $b \neq 0$, but $a = c = 0$, that, in addition to the triplet $(W^{\pm}, W^3)_L$ and the octet of gluons $V(8)$, the particular combination $B^0 \equiv (f_{15} W_R^3 - \sqrt{2/3}\, g V_{15})/(f_{15}^2 + \frac{2}{3} g^2)^{1/2}$ remains massless, the corresponding residual local symmetry thus being $SU(2)_L \times SU(3)'_{L+R} \times U(1)$.

[28]We may now point out the *one-to-one correspondence* referred to before in Sec. I between the scheme of spontaneous symmetry breaking proposed in Ref. 6 [see Eq. (17) and footnote 21 of this paper] and the one being considered by Mohapatra and Sidhu (Ref. 16). $(\kappa^2 + \kappa'^2)$ of Ref. 16 corresponds to a^2 of Ref. 6; likewise λ^2 of the former to $(b_4^2 + c_4^2)/2$ of the latter and zero of the former to $(b_4 - c_4)$ of the latter. The gauge coupling constants g and g' of Ref. 16 correspond to g and $(\frac{3}{2})^{1/2} f_{15}$, respectively, of Ref. 6. The question of whether the choice $b_4 - c_4 = 0$ can be maintained naturally [barring finite $\mathcal{O}(\alpha)$ corrections] in spite of $\langle E_L \rangle \neq \langle E_R \rangle$ within the context of a unified theory is being examined by H. S. Mani, J. C. Pati, and A. Salam.

[29]If $\epsilon_L \neq 0$, as stated earlier, all neutral-current amplitudes given in Secs. III and IV would simply be multiplied by a uniform enhancement factor $(c^2 + a^2 + \epsilon_L^2)/(c^2 + a^2) = (4 m_{W_L^+}^2)/(4 m_{W_L^+}^2 - g^2 \epsilon_L^2)$. Such a case offers the possibility [for a range of values of ζ, η, and r, see Eq. (28)] that one of the neutral gauge particles can be even lighter than the values presented in Sec. V (for $\epsilon_L = 0$). This possibility will be considered in more detail elsewhere.

[30]For many purposes it is convenient to use an alternative parametrization $b^2 \equiv c^2(1+x)$; $a^2 \equiv b^2 y$ with $-1 < x < \infty$, $0 \leq y < \infty$. Thus $x = (\eta - \zeta)/\zeta$ and $y = (1 - \zeta)/\eta$. See discussions in Sec. V for the usefulness of x and y.

[31]See for example, J. Bernabeu and C. Jarlskog, Nuovo Cimento 38A, 295 (1977).

[32]Note, of all the scalar multiplets introduced (A, B, C, D, E_L, E_R), only the multiplet A can have Yukawa coupling with the fermions.

[33]This dichotomy between neutrino versus electron-induced processes has its root once again simply in the fact that *available* neutrinos are essentially two component objects (ν_L's or $\bar{\nu}_R$'s), while electrons in atoms are four-component objects.

[34]V. Elias and A. R. Swift, Phys. Rev. D 13, 2083 (1976).

[35]It should be noted that case B is, of course, not necessarily exclusive of case A. One may satisfy the re-

quirements of both cases by setting $a^2=0$ (i.e., $\zeta=1$) and simultaneously setting $\eta=b^2/(c^2+a^2)\gg 1$.

[36]Note that the point $r=f_{15}^2/g^2=\frac{1}{3}$ corresponds [via Eq. (13a)] to $f_{15}^2/e^2=\frac{4}{3}$, which is the lower limit of the range $\frac{4}{3}e^2\leq f_{15}^2\leq 4e^2$ suggested by the renormalization-group considerations (Ref. 24).

[37]See for example, S. Sehgal, Phys. Lett. B48, 133 (1974); V. Barger and D. V. Nanopoulos, Nucl. Phys. B124, 426 (1977). The definitions of leptonic-scattering parameters $C_{\nu_\mu e}$, $C_{\bar{\nu}_\mu e}$, and $C_{\bar{\nu}_e e}$ follow those of B. W. Lee, Rapporteur talk at the Aachen International Conference, 1976 (unpublished). For example $\sigma_{\nu_\mu e}(10^{-41})\times E$) cm^2/GeV, similarly for $C_{\bar{\nu}_\mu e}$ and $C_{\bar{\nu}_e e}$.

[38]This can introduce an error ($\simeq$5–10%).

[39]Our inferences are, however, based on a much wider range of variation of the three relevant parameters $r=(f_{15}^2/g^2)$, $\zeta=c^2/(c^2+a^2)$, and $\eta=b^2/(c^2+a^2)$ than those exhibited in the tables and the figures.

[40]J. Blietschau *et al.*, CERN report, 1976 (unpublished); P. Musset, in *Proceedings of the XVIII International Conference on High Energy Physics, Tbilisi, 1976*, edited by N. N. Bogolubov *et al.* (JINR, Dubna, 1977), p. B87.

[41]L. Stutte, in *Proceedings of the International Conference on Production of Particles with New Quantum Numbers*, Madison, Wisconsin, 1976, edited by D. B. Cline and J. J. Kolonko (Univ. of Wisconsin, Madison, 1976), p. 388.

[42]A. Benvenuti *et al.*, report presented at the Conference on the Production of Particles with New Quantum Numbers, Madison, Wisconsin, 1976 (unpublished).

[43]H. Faissner *et al.*, reports at Aachen Conference (unpublished) and Tbilishi Conference (unpublished).

[44]F. Reines, H. Gurr, and H. Sobel, Phys. Rev. Lett. 37, 315 (1976).

[45]W. Lee *et al.*, Phys. Rev. Lett. 37, 186 (1976).

[46]D. Cline *et al.*, Phys. Rev. Lett. 37, 648 (1976).

[47]The comparison of theory with experiments for leptonic processes is still premature due to large errors in the experimental values. As regards elastic νp- and $\bar{\nu}p$-scattering parameters, the comparison is somewhat premature, since the theoretical predictions depend upon the knowledge of ν, $\bar{\nu}$ spectra and the mass characterizing the axial-vector form factor. We have followed D. P. Sidhu, Phys. Rev. D 14, 2235 (1976); C. Albright, C. Quigg, J. Smith, and R. Schrock, *ibid.* 14, 1780 (1976); R. M. Barnett, *ibid.* 14, 2990 (1976); and V. Barger and D. V. Nanopoulos (Ref. 37) in taking these features into account. We thank Professor Vernon Barger for most helpful communications in this regard.

[48]In Figs. 2, 3, and 4, we have exhibited the results for a fixed $r=(f_{15}^2/g^2)=0.44$. For a given $y=(1-\zeta)/\eta$, the neutral-current cross-section ratio $R=(\sigma_N^{\nu N}/\sigma_N^{\nu N}{}_C)$ increases with increasing r (see Table II). However, as long as the coupling-constant ratio r stays within the range 0.33 to 0.66 (the reasons for this restriction are discussed later), the cross-section ratio R decreases below the GGM range (0.59 ± 0.14) once y exceeds ≈ 0.28 (see Fig. 3). We should also remark that it is possible to obtain a fit for some of the parameters (for example, $R_{\bar{\nu}}$) with relatively large $y=(1-\zeta)/\zeta$, but then either the leptonic parameter $C_{\nu_\mu e}$ (Ref. 37) becomes too small compared to the present experimental range, or R becomes too small, or else R_ν becomes too large. [Note that in the SU(2)$\times$U(1) limit ($\eta\gg 1$), $y=(1-\zeta)/\eta$ vanishes; on the other hand, $x=(\eta-\zeta)/\zeta=(b^2-c^2)/c^2$ can be large (by $\zeta\to 0$) without y vanishing. Thus $x\to\infty$ is *not* necessarily the SU(2)$\times$U(1) limit; it is the SU(2)$\times$U(1) limit only provided $\zeta=c^2/(c^2+a^2)$ is nonzero.]

[49]The answer does not depend sensitively on ζ near $\zeta\approx 1$.

[50]While $m_{W_L^+}$ depends on r, the ratio $(m_{N_1}/m_{W_L^+})$ is not so sensitively dependent on r (see Table III).

[51]Allowing for $(g^2/4)(a_1^2+a_1a_4)\leq\frac{1}{4}m_{W_L^+}{}^2$ and $m_{W_R^+}{}^2\gtrsim 10m_{W_L^+}{}^2$, the lowering of W_L^+ mass due to $W_L^+-W_R$ mixing [see (9b)] is given by

$$(g^2/4)\frac{(a_1^2+a_1a_4)^2}{a^2+b^2+\epsilon_R^2}=(\lesssim\tfrac{1}{4}m_{W_L^+}{}^2)\left(\frac{\lesssim\frac{1}{4}m_{W_L^+}{}^2}{\gtrsim 10m_{W_L^+}{}^2}\right)$$

$$\lesssim(m_{W_L^+}{}^2/160),$$

which is a small effect. Likewise the $W_L^+-W_R^+$ mixing angle is given by

$$2(a_1^2+a_1a_4)/(a^2+b^2+\epsilon_R^2-a^2-c^2-\epsilon_L^2)$$

$$=2\left(\frac{\lesssim\frac{1}{4}m_{W_L^+}{}^2}{\gtrsim 10m_{W_L^+}{}^2}\right)\lesssim\tfrac{1}{10},$$

which also is a small effect for most purposes.

International Journal of Modern Physics A Vol. 1 No. 1 (1986) 147–153
© World Scientific Publishing Company

THE MOON AS THE SEARCHING GROUND FOR PROTON DECAY*

J. C. PATI
Department of Physics and Astronomy
University of Maryland, College Park, MD 20742, USA

ABDUS SALAM
International Centre for Theoretical Physics
34100 Trieste, Italy

and

B. V. SREEKANTAN†
Department of Physics and Astronomy
University of Maryland, College Park, MD 20742, USA

Received 7 November 1985

The distinct advantage which a lunar detector would have over a terrestrial one in searching for proton decays corresponding to rather long nucleon lifetimes $\gtrsim 6 \times 10^{32}$ yrs is noted.

Proton-decay searches at various laboratories of the world have so far led to results which are either negative or whose interpretations are subject to some ambiguity. The IMB group[1] working with the largest detector, particularly well-designed for detecting the $e^+\pi^0$-mode, have as yet reported no candidates for this mode, setting a lower limit $[\tau_p/B(p \to e^+\pi^0)] \gtrsim 2 \times 10^{32}$ yrs. On the other hand, several groups[2–5] have reported contained events which could arise from genuine proton decays (in the experimenters' views) of the type $p \to \mu^+ K^0$, $p \to \mu^+ \eta$, etc. Unfortunately the interpretations of these events appear to be subject to ambiguity, especially if one demands near consistency between the observations of the different groups. The sole cause of this ambiguity is the expected background owing to atmospheric neutrino-induced reactions, whose rate in the relevant energy-range roughly appears to coincide with that of the observed events.

Since proton decay bears on fundamental issues,[6] one would, of course, like to see that the efforts in its search be strengthened in *all directions* so that one may finally ascertain beyond a shadow of doubt whether or not it decays with observable lifetimes $\lesssim 10^{34}$ yrs (say).*

If the background-to-signal ratio in the search for proton decay continues to be a problem even after the present detectors have run for the next three to four years (whether or not this will happen largely depends, of course, upon the proton lifetime and the nature of its dominant decay modes), it seems to us that one must then consider

† Permanent address: Tata Institute of Fundamental Research, Bombay, India

* From a theoretical standpoint, within grand unified symmetries like SO(10) and SU(16) which unify members of one family in contrast to SU(5) and which permit a *two-stage-descent* into $SU(2)_L \times U(1) \times SU(3)^c$, proton lifetime is expected to be of order 10^{31}–10^{34} yrs for such a two-step descent with $\sin^2\theta_W \simeq 0.22$–$0.23$.

seriously the possibility of using the moon as the searching ground for proton decay. The major advantage of the moon in this regard is the fact that it has no atmosphere and, therefore, no atmospheric neutrinos or atmospheric muons. The background of neutrinos, muons and neutrons arising from interactions of primary cosmic rays with the moon rock and/or the detector materials can be made sufficiently small by a suitable design and a choice of the location of the detector, as we discuss, so that even a few events of the type observed on the earth (e.g., the $\mu^+ K^0$-type), if seen on the moon, should constitute evidence for proton decay. As we shall see, the galactic and extragalactic neutrino backgrounds do not pose a serious problem on the moon either.

The purpose of this note is to stimulate thinking in this regard. Conscious of our lack of expertise in rocket technology and its limitations and also of our knowledge of detectors which may be available, say five to ten years from now, and the lack of detailed information about the surface and the underground structures of the moon, we still wish to make a start here by spelling out what appear to us to be some of the relevant issues.

We are encouraged to entertain the idea of moon experiments in the first place because several space planners are now enthusiastic to revive trips to the moon and have suggested that few people may even live there, mining raw materials, growing their own food and using solar energy.[b] This could provide a very useful experience which is needed for exploration and possible colonization of space in the future. If such an exploration of the moon materializes, it would become much easier to entertain the idea of setting up the proton-decay experiment on the moon as a parasitic endeavor to the main program at a very small fraction of the total cost. We therefore proceed with the presumption that the moon exploration will resume in full force in the not too distant future.

Since we wish to have a source which effectively provides at least (few) $\times$ (10^{33} to 10^{34}) nucleons, which would need at least (few) $\times$ (10,000 tons) of matter,[c] it is clear that we must utilize the moon rock (or dust) itself to provide the bulk of this. In other words, we should plan to carry from the earth only those components and systems which are needed to detect the decay products, and to record and transmit the information.

Since the flux of primary cosmic rays unhindered by the atmosphere and magnetic field would be rather enormous at the surface of the moon (e.g., the flux of 10 GeV protons is $\simeq$ 1 per cm^2 per sec $\sim 10^{14}$ per 1000 m^2 per year) we must reduce this and also the secondaries produced in these interactions, especially the neutrons, by at least 15 to 16 orders of magnitude (say) to help isolate proton-decay signals which may be of the order of 10 to 5 per year for a 20 kiloton detector with $\tau_p \sim$ (1 to 2) $\times 10^{33}$ yrs (say). To reduce the flux by this extent, we must shield the detector by placing it, for example, in a tunnel with, say, 100 meters of rock above it (see Fig. 1).[d] One may, for

[b] For a descriptive report on this, see e.g. J. Bigger, Washington Post Magazine Section, p. 8, September 1, 1985.

[c] 10,000 tons of matter yields about 6×10^{33} nucleons.

[d] By having at least 100 meters of rock above the detector, the flux of primary cosmic rays would be cut down by a factor of at least 10^{20}. See discussions later about possible background due to secondary particles.

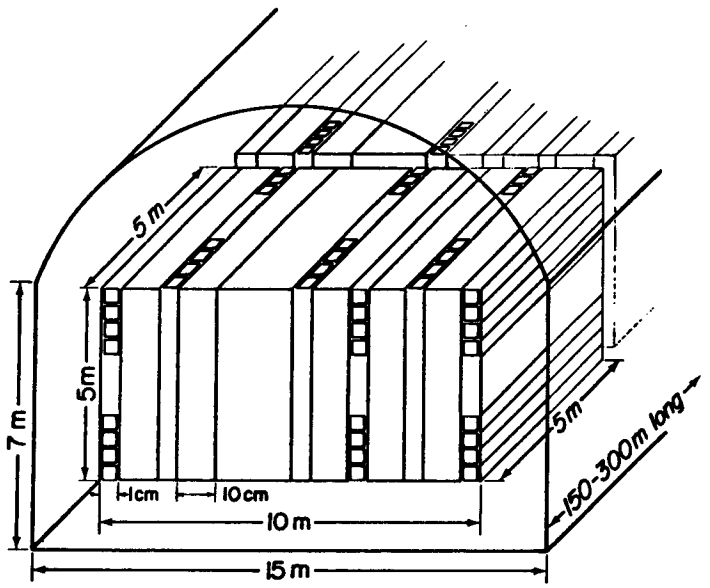


Fig. 1

example, blast a tunnel on one side at the bottom of a crater which is fairly deep[e] (depth $\gtrsim$ 100 meters, say) but not too wide at the top.

While the exact details of the nucleon decay detector to be installed on the moon will naturally depend on the state-of-the-art of detector technology, the limitations on the size and weight of individual packages and the restrictions on the type of gases that can be carried to the moon, etc., we may for purposes of illustration envisage systems that are similar to the ones operating in the terrestrial underground laboratories. Since there appear to be no water or any other transparent liquid readily available on the moon, we have to rule out water Cerenkov type of detectors and plan in terms of the fine grain calorimeters. In line with the currently operating systems the detector could be modular in construction, *each module* having the following features: overall dimensions $\approx$ 5 m (height) $\times$ 5 m (length) $\times$ 10 m (width) (see Fig. 1), composed of 80 vertical layers of moon rock packed in suitable lightweight containers—each 5 m $\times$ 5 m $\times$ 10 cm, interspersed with 80 crossed layers of gaseous discharge counters made of lightweight plastic walls. Each counter in the layer may be 5 meters long with a cross-sectional area of 1 cm $\times$ 1 cm. The lunar rock (with a packing density[f] of about 2 g/cm^3) in such a module would weigh approximately 400 tons and the plastic containers nearly 2 tons.

For a 10–20 kiloton nucleon detector, we require 25–50 such modules and the length

[e] If the idea of mining raw materials on the moon materializes, then blasting a tunnel may not need new equipment.

[f] The average density of the moon is $\simeq$3.39 g/cm^3. That at the surface is $\approx$2.2–2.8 g/cm^3, corresponding to rocks like basalt.

of the tunnel for the detector system alone would be 125–250 m—the overall dimensions of the tunnel would have to be approximately 150–300 m long, 15 m wide and 7 m high. From the mouth side of the tunnel there would be a background of neutrons produced by the bombardment of the unhindered cosmic rays on the surface of the crater. To reduce this background, the assembling of detector modules should start sufficiently deep inside the tunnel—i.e. about 30–40 meters from the mouth—and this space could be utilized for all the electronic and other equipment.

We see that for a 10–20 kiloton detector, the total earth-weight of the detector equipment which needs to be carried from the earth to the moon for the sake of the proton-decay experiment is around 25–50 modules × 2 tons = 50–100 tons. The recording and transmitting equipment may weigh at least half as much. Once the moon exploration project is established, the cost of carrying this weight from the earth to the moon might be manageable.[8]

Expected Background

On the moon one does not have to worry about atmospheric neutrinos or muons. But secondary hadrons like $\pi^{\pm}$ and kaons produced by interactions of the primary cosmic rays with the moon rock or dust can decay into muons and neutrinos. (Neutrinos can also arise from the secondary muon decays.) These will provide a source of background for nucleon decays, provided $E_{\nu} \sim 1\,\text{GeV} \pm 200\,\text{MeV}$. Fortunately, owing to the intervening moon rock or dust, which is much more dense ($\rho \gtrsim 2\ \text{g/cm}^3$) than the earth's atmosphere, the secondary hadrons will most of the time interact strongly with the lunar rock, lose energy and be absorbed before having a chance to decay. For example, for a 2 GeV pion, the effective decay length is about 110 meters, whereas the interaction mean free path is less than 1/2 m. For this reason, the neutrino background ($E_{\nu} \sim 1$ to 2 GeV) arising from secondary decays in the moon would be reduced by at least a factor of 200 compared to that on the earth. This is the major advantage of the moon experiment as stressed before.

The production of energetic muons ($E_{\mu} \gtrsim 2$ GeV) arising from $\pi^{\pm}$ (or $K^{\pm}$) decays would also be considerably reduced compared to that in the earth's atmosphere for the same reason as above—i.e., most of the correspondingly energetic $\pi^{\pm}$ (or $K^{\pm}$) will not have a chance to decay. The higher the energy of pions (or kaons), the lower is the relative probability for their decays. Those muons which would still arise from pion (or kaon) decays with energies ~ 1 to 5 GeV (say) would rapidly lose energy in the rock due to ionization loss; the range of a 5 GeV muon in the moon rock with a density $\simeq 2.5\ \text{g/cm}^3$ is about 10 meters, while the decay length is nearly 9 km. Thus the relevant muon flux arriving at the nucleon decay detector which is shielded by moon rock ($\gtrsim 100$ m) on all sides would be reduced by at least five to six orders of magnitudes compared to that on the earth. For this reason, there would be no

[8] Any cost estimate would, of course, be significantly lowered beyond the initial stages of the exploration of the moon.

equilibrium neutron background, either. Furthermore, the component of the neutrino background arising from muon decays would be essentially eliminated. This would further reduce the net neutrino background on the moon relative to the earth. Conservatively, we will still take the relative reduction factor to be $\gtrsim 200$.

So far, we have not considered the contribution from galactic neutrinos which are predominantly produced in the interaction of cosmic rays with the interstellar gas and dust. Due to the lens shape of the galaxy and due to the fact that the solar system is located off the center, this flux is highly anisotropic as seen on the earth and on the moon. The maximum flux will naturally be from the direction of the galactic center and the minimum from a direction perpendicular to the galactic plane. Detailed calculations[7,8,9] show that (i) the neutrino flux from the direction of the galactic center (i.e., from a band distributed along galactic longitudes of -60 to $+60°$ and latitudes of $\pm 6°$) is lower by a factor of nearly one hundred compared to the terrestrial atmospheric neutrino flux; (ii) it is further reduced by a factor of 5–10 from other directions in the galactic plane and by a factor of about 100 in a direction perpendicular to the galactic plane compared to that from the galactic center. Therefore it is clear that the galactic neutrino flux in the lunar underground laboratory would be comparable to the locally produced neutrino flux discussed above only in a restricted solid angle in the direction of the galactic center, or at the worst in all azimuths within the galactic belt of say $\pm 20°$. However, if we restrict ourselves to events away from this solid angle, the galactic neutrinos are again down by at least three orders of magnitudes[h] compared to the atmospheric neutrino background on the earth.

To summarize, the backgrounds due to all sources—i.e., neutrinos, muons and neutrons—can be reduced by at least a factor of two hundred and perhaps even better on the moon compared to those on the earth.

In Table 1, we present a comparative summary that exhibits the distinct advantage which a lunar detector would have over a terrestrial one. For example, if $\tau_p \gtrsim 10^{33}$ yrs, the expected ν background event rate on the earth, even after the relevant reduction due to topological cuts, would be equal to or *exceed* the rate of proton decay events.[i] On the other hand, the background-to-signal ratio on the moon, with the same topological reduction factor as on the earth, is at best only 0.5% for nucleon lifetimes (τ_N) of nearly 10^{33} yrs and 5% for $\tau_N \simeq 10^{34}$ yrs.

[h] The galactic and extragalactic neutrino flux calculations depend on (i) the assumed distribution of matter and (ii) distribution of cosmic rays in the galaxy and outside. If these are radically different from what is assumed in the calculations, then the same experiment on the moon will serve to open up new vistas in our understanding of the structure of matter and radiation in the galaxy.

[i] For this reason, 10^{33} yrs seems to be about the limit which proton-decay searches on the earth can optimistically hope to reach as regards proton lifetime, at least under the present state-of-the-art. Since the raw neutrino background cannot be reduced, the only way to reduce the effective background is to improve the reduction factor beyond that of 100 shown in Table 1. If detectors are designed so as to improve this factor significantly *and such efforts should by all means be undertaken*, it is clear that such improved detectors transplanted to the moon would, in principle, permit searches for proton-decays with lifetimes approaching even 10^{35} yrs.

Table 1

τ_N	Nucleon Decays Expected in 20 KTY	ν-Induced Events for Earth Detector $E_\nu = 1000 \pm 200$ MeV: Without topological cut	ν-Induced Events for Earth Detector $E_\nu = 1000 \pm 200$ MeV: With topological cut	ν-Induced Events for Moon Detector $E_\nu = 1000 \pm 200$ MeV: Without topological cut	ν-Induced Events for Moon Detector $E_\nu = 1000 \pm 200$ MeV: With topological cut
10^{32} years	120	1000	10	<5	<0.05
10^{33} years	12	1000	10	<5	<0.05
10^{34} years	1.2	1000	10	<5	<0.05

A comparative summary of earth-versus-moon nucleon decay detectors. The fourth and the sixth columns show the expected number of events due to neutrino interactions which would simulate nucleon-decay events in the earth and the moon detectors respectively for 20 KTY exposure in each case. Based on observations and theoretical calculations, we have assumed that for a detector on the earth, there will effectively be a total of nearly 50 neutrino induced events with $E_\nu \simeq (1000 \pm 200)$ MeV per KTY-exposure and that this background can be further reduced by about a factor of a hundred for high resolution fine grain detectors by imposing topological cuts (see e.g., Ref. 10). The raw neutrino background on the moon is expected to be lower by at least a factor of 200 than that on the earth (see text). It is assumed that the background reduction factor due to topological cuts is the same for the moon as it is for the earth detector.

Thus, if the technical and financial aspects involving transportation of detector materials weighing about 100 to 200 tons from the earth to the moon and setting up the detector on the moon can be managed, proton decays with lifetimes $\lesssim 10^{34}$ yrs can be measured with confidence on the moon. This would be a significant improvement over terrestrial experiments. Needless to say, the proton-decay detector set up on the moon will also be most useful as a detector of neutrinos from astronomical sources because of the significant reduction of the background due to local sources (like the atmosphere).[j] Furthermore, it would permit one to search for neutron-antineutron oscillations with far greater sensitivity than that available on the earth.

We believe that many physicists must have already given thought to the idea of searching for proton decay on the moon.[k] We, therefore, hope that this note will help focus on the idea and generate further thinking leading to an optimum design of a proton-decay detector specially suited for the moon.

Acknowledgments

We thank S. Miyake, George A. Snow, John Strathdee and G. B. Yodh for many helpful discussions. The research of JCP is supported in part by the National Science Foundation.

[j] This has been noted e.g. by M. Shapiro and R. Silberberg, M. Cherry and K. Lande and A. G. Petschek in abstracts of the *Symposium on Lunar Bases and Space Activities in the 21st Century*, held Oct. 29–31, 1984, published by L. B. Johnson Space Center, Houston, Texas.

[k] In the process of writing this note, we came across an abstract by A. G. Petschek (see footnote j) mentioning this point. We thank Dr. C. Alley for kindly drawing our attention to the abstracts. For a short remark, see also M. Goldhaber, P. Langacker and R. Slansky, *Science* **210** (1980) 851. We have subsequently learned that F. Reines has also been considering this idea (private communication).

References

1. R. Bionta *et al.*, *Phys. Rev. Lett.* **51** (1983) 27; *Proc. 22nd Int. Conf. on High Energy Phys.*, Vol. I Leipzig (July 1984) p. 244.
2. M. R. Krishnaswamy *et al.*, *Pramana* **19** (1982) 525; *Proc. 22nd Int. Conf. on High Energy Phys.*, Vol. I Leipzig (July 1984) p. 244.
3. G. Battistoni *et al.*, *Phys. Lett.* **118B** (1982) 461; *Proc. 22nd Int. Conf. on High Energy Phys.*, Vol. I Leipzig (July 1984) p. 246.
4. M. Koshiba *et al.*, *Proc. 22nd Int. Conf. on High Energy Phys.*, Vol. I Leipzig (July 1984) p. 250.
5. E. Aprile-Giboni *et al.*, *Proc. 22nd Int. Conf. on High Energy Phys.*, Vol. I Leipzig (July 1984) p. 252.
6. J. C. Pati and A. Salam, *Phys. Rev.* **D8** (1973) 1240; *Phys. Rev. Lett.* **31** (1973) 661; H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **32** (1974) 438.
7. V. S. Berezinsky, *Proc. 19th Int. Conf. on Cosmic Rays*, Vol. **6** Budapest (1977) p. 231.
8. R. Silberberg and M. M. Shapiro, *ibid.*, (1977) p. 237.
9. A. Dar, *Proc. 4th Workshop on Grand Unification*, p. 101 (1983).
10. A. Grant, *Proc. 1982 Summer Workshop on Proton-Decay Experiments*, Argonne National Lab., ANL-HEP-PR-82-24, ed. by D. S. Ayres, p. 203.

4. Gravity, Supersymmetry and Strings

The main focus of Abdus Salam's research work in theoretical physics has always been directed towards quantum field theory and the unification of the fundamental forces. Therefore, it is not surprising that at various times he has extended his activities to include the gravitational field.

The problem of constructing a coherent and consistent theory of quantum gravity has proved remarkably intractable and, even today, a variety of views are held on the subject. From the perspective of an elementary particle physicist it is natural to start by reducing general relativity to something that resembles a standard, Poincaré-invariant theory. This can be done by writing the space-time metric tensor $g_{\mu\nu}(x)$ as $\eta_{\mu\nu}+h_{\mu\nu}(x)$ where $\eta_{\mu\nu}$ is the usual Minkowskian metric on flat space-time, and then regarding $h_{\mu\nu}(x)$ as a field that is to be quantized in a standard way using (special-)relativistic quantum field theory. This works well at the linearized level and yields a theory of spin-two, massless gravitons propagating on the background Minkowski space-time. However, when substituted into the Einstein-Hilbert action, the expression $\eta_{\mu\nu} + h_{\mu\nu}(x)$ gives a Lagrangian for the graviton field $h_{\mu\nu}(x)$ that is non-polynomial and, by elementary power counting, the ensuing quantum field theory is highly non-renormalizable. Two traditional reactions to this situation have been to suggest that (i) *non*-perturbative quantization techniques are required; or (ii) one must look for a modification of general relativity in which the unwanted ultraviolet divergences somehow cancel each other.

Salam's interest in quantum gravity was kindled in the late 1960s by his work on the quantization of non-polynomial Lagrangians (see Sec. 1). He realised that it should be possible to apply these techniques to the case of quantum gravity and began a detailed research programme on the topic. The paper with Isham and Strathdee illustrates the type of result that was obtained when the methods were applied to gravity-modified quantum electrodynamics: for example, the gravitational self mass δm of the electron is calculated to be proportional to $m\alpha \log(M_{\rm P}/m)$ where $M_{\rm P}$ is the Planck mass. Another early paper on quantum gravity is the article with Delbourgo in which the gravitational correction to the PCAC rule is calculated. This seemed rather esoteric at the time but, in fact, it anticipated by many years the extensive later interest in gravitational contributions to anomalies in the context of superstring theories.

Although interesting, these non-perturbative approaches had problems of their own and did not find wide acceptance. Thus, in the early 1970s, research in quantum gravity was divided mainly between a group of quantum field theorists — who were rather gloomily confirming in detail that the perturbative quantization of

general relativity really *is* non-renormalizable — and a number of general relativists who were tackling the problem from a quite different, canonical approach, and concentrating on applications to quantum cosmology and the difficult conceptual problems this raises. It is worth noting that, prior to this era, there had been very little interaction between the particle physics and general relativity communities, and one of the first meetings where the two schools came together was organized in Trieste by Salam in 1970. Many of the contacts forged then have continued to be of value to the present day.

The particle-physicists' contribution to quantum gravity could well have become rather moribund at this time had it not been for the timely discovery of *supersymmetry*. In theories of this type, fermions and bosons are paired together in a fundamental way, which has the great attraction that, since fermion loops in quantum field theory have the opposite sign to bosonic loops, the possibility arises of ill-behaved bosonic theories (like quantum gravity) being 'tamed' by forcing them to become supersymmetric.

Working with his long-term associate John Strathdee, Salam produced a number of papers in which techniques for handling theories of this type were developed. Of particular importance are those dealing with 'superfields' in which the many fields needed to constitute a supermultiplet that transforms covariantly under a supergroup are combined in an elegant way to form a single mathematical entity $\Psi(\theta, x)$. The variable θ is of Grassmann type, and hence expanding Ψ in θ leads to a formal power series that truncates at a finite order; the component fields in the expansion are then precisely the components of the supermultiplet.

The importance of this technique can be illustrated by considering the early ways of constructing supersymmetric theories. Typically, this involved starting with a bosonic Lagrangian to which terms were carefully added by hand until a supersymmetric result is obtained — a procedure that is rather *ad hoc* and often complex and time consuming to implement. On the other hand, the use of superfields provides a simple way of constructing classical field systems that are manifestly invariant under supersymmetric transformations. Salam and Strathdee also showed that it was possible to quantize superfields directly to give Feynman rules for theories of this type which, if applied in the appropriate way, can be guaranteed to yield theories that are manifestly supersymmetric at the quantum level.

Another significant paper from this era is the work on unitary representations of supersymmetry transformations in which the well-known induced-representation methods of Wigner and Mackey are applied to the supersymmetry group, thereby providing a powerful way of exploring the physical particle content of theories of this type.

The obvious implications of supersymmetry for quantum gravity were quickly realised and a number of supergravity theories were constructed in the mid and late 1970s. The hope was that one or more of these theories would be completely free of ultraviolet divergences and would thus solve the problem of quantum gravity. In addition, since it would involve a precise mix of bosonic and fermionic fields, the ensuing framework might constitute a genuine 'unified' theory in which the graviton is combined in a natural way with the quanta of the other fundamental forces of nature.

Broadly speaking, the more supersymmetry generators there are, the better are the chances of removing ultraviolet infinities, and hence there was great interest in studying the largest theory that is possible, i.e., $N = 8$ supergravity. However, this theory has the striking feature of being most naturally formulated in a space-time of dimension *eleven* and this reinforced the concurrent interest in studying Kaluza-Klein type theories in which the basic fields are defined on a fibre bundle that sits over standard four-dimensional space-time. A single field on the big space reduces to an infinite set of fields on the base space-time, and the calculations of the resulting particle spectra and Lagrangians are frequently extremely complex. The paper 'On Kaluza-Klein Theory' by Salam and Strathdee was a major contribution to understanding the full structure of such theories in the important case where the extra dimensions (i.e., the 'fibre' of the bundle) form a homogeneous space G/H where G and H are Lie groups. This study involved an extensive spectral analysis over this space and linked back nicely to their earlier work on the use of nonlinear group realizations in a variety of physical situations.

One of the major challenges in all field theories defined on higher-dimensional space-times is to construct solutions to the classical field equations that can serve as ground states for the quantum theory in which some sort of spontaneous compactification occurs to yield a stable theory on a physically acceptable four-dimensional space-time. Several papers in this collection are concerned with a particularly elegant example of this process in the context of a gauge theory on a six-dimensional space in which a stable compactification over the two-sphere fibre is achieved with the aid of a non-trivial configuration (a monopole) of the electromagnetic field on the two-sphere. This is a nice example of the use of topological ideas in quantum field theory. It also illustrates one way of producing the chiral fermions that are necessary for a physically-relevant theory of fundamental forces but which, as Witten first emphasized, can be rather elusive in theories of this type.

A major international effort was devoted to the development of supergravity theories and it was shown that, as expected, some of the troublesome ultraviolet divergences of the conventional theory are indeed ameliorated. However, these theories were also shown to admit potential supersymmetric counterterms, and thus

higher-loop divergences could still appear. This has never been checked explicitly because of the great complexity of the necessary calculations, but it is now generally agreed that supergravity by itself does *not* solve the problem of quantum gravity.

The response of general-relativists working in quantum gravity has mainly been to explore further the idea of non-perturbative quantization, of which the most developed scheme is the canonical approach based on the new variables of Abhay Ashtekar. However, the particle physics community has moved in a different direction and has concentrated its attention on the study of superstring theory in which the paradigmatic point particles of supergravity are replaced with tiny loops propagating in space-time. Supergravity still appears as the low-energy limit of the theory but the Planck-energy domain is radically different and leads to major new results. Salam himself was not directly involved in the discovery of these theories but it is a topic in which he has shown considerable interest and the one with which the final two papers in this selection are concerned.

The great interest in superstring theory as a possible theory of gravity started with the construction by Green and Schwarz in the early 1980s of a version that is anomaly free and with a good chance of having no ultraviolet divergences. However, this particular theory worked only in a space-time of 10 dimensions, and a variety of methods have been explored for constructing theories directly in lower dimensions. The paper by Salam and Bergshoeff *et al.* is a nice example, being one of the first works in which the critical dimension is computed for a theory where the target space for the strings is a Lie group manifold. The paper on σ-models and string theories is also of note for the way in which the requirement of quantum consistency (freedom from conformal anomalies) is related to global properties of the two-dimensional space traced out by the moving string: a topic that has generated considerable interest in recent years.

It is noteworthy that almost all of Salam's papers on gravitational matters were written with John Strathdee, a major theoretical physicist in his own right whose knowledge and skill in the application of mathematical techniques are formidable. The complementary technical abilities and creative skills of the two men produced an exceptionally productive long-term collaboration whose fruits are particularly well illustrated by the papers in this section.

Reprinted from THE PHYSICAL REVIEW D, Vol. 3, No. 8, 1805–1817, 15 April 1971

Infinity Suppression in Gravity-Modified Quantum Electrodynamics

C. J. ISHAM, ABDUS SALAM,* AND J. STRATHDEE
International Centre for Theoretical Physics, Miramare-Trieste, Italy
(Received 28 September 1970)

Computations of gravity-modified quantum electrodynamics are performed using nonpolynominal Lagrangian field-theory techniques. The inverse of the gravitational constant appears as an effective cutoff mass, and, in particular, it is shown that to order e^2 the electron and photon self-energies are finite. The cutoff can be interpreted as if the electron had an intrinsic radius equal to its Schwarzschild radius. A central feature is the construction of the tensor gravity superpropagator.

I. INTRODUCTION

IT has been conjectured in the past that the universal and nonlinear coupling of gravitation to matter may provide a natural mechanism for the damping of ultraviolet infinities in field theory.[1] In a recent Letter[2] we revived this conjecture and pointed out that the newly developed techniques for computing with nonpolynomial Lagrangians would lend themselves to testing it. It is the purpose of this paper to show by actual computations to second order in the electromagnetic coupling that the conjectured damping indeed happens in gravity-modified electrodynamics. Our results are

$$\frac{\delta m}{m} \approx \frac{3\alpha}{4\pi} \ln\left(\frac{1}{\kappa m}\right)^2 \approx \frac{2}{11},$$

$$\frac{\delta e}{e} \approx -\frac{\alpha}{6\pi} \ln\left(\frac{1}{\kappa m}\right)^2 \approx -\frac{1}{25},$$

where κ (the square root of 16π times the Newtonian constant G) equals 0.5×10^{-18} GeV^{-1}. These results are of interest because in spite of the extreme smallness of the gravitational constant κ, the values obtained for $\delta m/m$ and $\delta e/e$ are a reasonable order of magnitude.[3,4] It is amusing that the effective cutoff appears to come at a length which equals the Schwarzschild radius of the electron $R_c = 2m_e{}^2G$ (measured in units of m_e^{-1}). As pointed out recently,[4] it seems not unreasonable to hope that when higher-order terms in the effective perturbation parameter $\alpha \ln(\kappa m)^2$ are included, one will find, for example, that nearly all of the electron's mass can be explained as electromagnetic ($\delta m/m \approx 1$).

This paper is in the nature of a report on how these results were arrived at and is planned as follows. In Sec. II we review Einstein's gravity theory—particularly as applied to electrons—and state the Feynman rules for graviton Lagrangians. The infinity-suppression mechanism is presented in Sec. III, which is devoted to the construction of the graviton superpropagator, and in Sec. IV, where the superpropagator is used in the explicit computation of the leading parts of the above-mentioned quantities $\delta m/m$ and $\delta e/e$. In Sec. V we briefly discuss the question of gauge invariance and equivalence theorems. The details of the computations needed for Sec. IV are contained in the Appendix. Apart from the numerical results, the item of greatest interest in the paper may be the graviton superpropagator exhibited in Sec. III.

II. GRAVITATIONAL LAGRANGIAN AND FREE PROPAGATOR

Einstein's theory of gravity is based upon the requirement that the equations of physics should be expressible in a form which is independent of the system of space-time coordinates to which they are referred. That is, these equations should be covariant under the group of general coordinate transformations; they should be derived by varying an action integral which is itself invariant under this group. To formulate the requirement one must first assign the variables of the problem to realizations of the transformation group. The following remarks are intended to sketch the main features of this program together with the steps necessary for treating quantum gravity.

The basic field upon which the group of general coordinate transformations are realized is the metric tensor $g_{\mu\nu}$ or, equivalently, the vierbein system $L_{\mu a}$. On the one hand, the metric tensor transforms according to the usual prescription

$$g_{\mu\nu}(x) \to \bar{g}_{\mu\nu}(\bar{x}) = \frac{\partial x^\alpha}{\partial \bar{x}^\mu}\frac{\partial x^\beta}{\partial \bar{x}^\nu} g_{\alpha\beta}(x), \qquad (2.1)$$

where $\partial\bar{x}^\mu/\partial x^\alpha$ denotes the Jacobian matrix of the transformation $x^\mu \to \bar{x}^\mu$. The vierbein components, on the other hand, transform according to a hybrid prescription

$$L_{\mu a}(x) \to \bar{L}_{\mu a}(\bar{x}) = \frac{\partial x^\nu}{\partial \bar{x}^\mu} \Lambda_a{}^b L_{\nu b}(x), \qquad (2.2)$$

* International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England.

[1] See, for example, S. Deser, in Proceedings of the Symposium on the Last Decade in Particle Theory, Center for Particle Theory, University of Texas, Austin, 1970 (unpublished).

[2] R. Delbourgo, Abdus Salam, and J. Strathdee, Nuovo Cimento Letters 2, 354 (1969).

[3] Similar results have been obtained by F. Hoyle and V. Narlikar [Ann. Phys. (N. Y.) (to be published)] using different techniques, and by I. B. Khriplovich [Yadern. Fiz. 3, 575 (1966) [Soviet J. Nucl. Phys. 3, 415 (1966)]] using an integral equation for the mass operator. We are indebted to H. Pegis for bringing this work to our attention. See also B. S. DeWitt, Phys. Rev. Letters 13, 114 (1964).

[4] Abdus Salam and J. Strathdee, Nuovo Cimento Letters 4, 101 (1970).

where $\Lambda_a{}^b$ denotes a Lorentz matrix,

$$\Lambda_a{}^b\Lambda_c{}^d\eta_{bd}=\eta_{ac}\,, \tag{2.3}$$

with $\eta_{ab}=\text{diag}\,(+1,-1,-1,-1)$. The metric tensor field can be expressed in terms of the vierbein field by means of the formula

$$g_{\mu\nu}=L_{\mu a}L_{\nu b}\eta^{ab}\,. \tag{2.4}$$

Conversely, the vierbein field can be expressed in a (locally) unique fashion in terms of the metric if the 16 components $L_{\mu a}$ are reduced to 10 independent ones by imposing the symmetry condition

$$L_{\mu a}=L_{a\mu}\,. \tag{2.5}$$

(In a pseudo-Euclidean notation where η_{ab} is effectively replaced by the Kronecker symbol δ_{ab}, the symmetric matrix L is given by a square root of the symmetric matrix g.)

It is particularly important to realize that the symmetry condition (2.5) not only fixes L in terms of g, but in addition serves to determine the Lorentz matrix Λ of (2.2) as a function of $\partial\bar{x}/\partial x$ and L itself. This results from the (local) uniqueness of the polar decomposition of the matrix $F_{\mu a}=\partial x^\nu/\partial\bar{x}^\mu L_{\nu a}$ into the product of a symmetric matrix $\bar{L}_{\mu b}$ and a (pseudo) orthogonal matrix $(\Lambda^{-1})_a{}^b$. Thus, corresponding to each coordinate transformation $x^\mu\to\bar{x}^\mu$, one has an associated Lorentz transformation Λ of the vierbein system,

$$\Lambda_a{}^b=\Lambda_a{}^b(\partial\bar{x}/\partial x,g)\,.$$

The function indicated here is generally nonlinear. One has constructed in this way a nonlinear realization of the group of general coordinate transformations.[5]

Any local field which belongs to a representation of the Lorentz group can be made to carry a nonlinear realization of the general group. Thus

$$\psi(x)\to\bar{\psi}(\bar{x})=\left|\det\frac{\partial\bar{x}}{\partial x}\right|^w D(\Lambda)\psi(x)\,, \tag{2.6}$$

where Λ denotes the Lorentz transformation determined above and w denotes a new parameter, the *weight*, which must be assigned. (For the subgroup of Lorentz transformations on space-time we have $\Lambda=\partial\bar{x}/\partial x$ and $|\det\partial\bar{x}/\partial x|=1$, so that w becomes irrelevant.) Nonlinear realizations based on Lorentz four-vectors and their products can be made over into linear representations of the conventional sort with the help of the vierbein. Thus, for example, if

$$\phi_a\to\bar{\phi}_a=\Lambda_a{}^b\phi_b\,,$$

then the combination $\phi_\mu=L_{\mu a}\eta^{ab}\phi_b$ transforms according to

$$\phi_\mu\to\bar{\phi}_\mu=\frac{\partial x^\nu}{\partial\bar{x}^\mu}\phi_\nu\,.$$

The spinorial nonlinear realizations cannot be linearized in this way. In order to treat the gravitational interactions of fermions it is essential to construct the vierbein components.

The scheme adopted here for introducing gravity by means of a nonlinear realization technique, which is well known to differential geometers, is not the usual one adopted in the physics literature. In the latter it is assumed that the vierbein field has 16 independent components $L_{\mu a}$ which transform under the coordinate group according to

$$L_{\mu a}\to\frac{\partial x^\nu}{\partial\bar{x}^\mu}L_{\nu a}$$

and under an *independent* "gauge group" according to

$$L_{\mu a}\to\Lambda_a{}^bL_{\mu b}\,,$$

where $\Lambda_a{}^b$ is not related to $\partial\bar{x}/\partial x$. In this view the field of, say, a Dirac particle would comprise a true Dirac spinor under the gauge group and a set of scalars under the coordinate group. This view is of course equivalent to the one we adopt.

In order to obtain an invariant action integral, one must construct the Lagrangian function so that it transforms as a scalar density,

$$d^4x\,\mathfrak{L}(\psi)=d^4\bar{x}\,\mathfrak{L}(\bar{\psi})\,,$$

that is,

$$\mathfrak{L}(\psi)\to\mathfrak{L}(\bar{\psi})=\left|\det\frac{\partial x}{\partial\bar{x}}\right|\mathfrak{L}(\psi)\,. \tag{2.7}$$

When expressed in terms of fields which belong to nonlinear realizations, $\mathfrak{L}(\psi)$ clearly must take the form of a Lorentz scalar with $w=-1$. Such a Lagrangian can be generated from any Lorentz invariant one by introducing $L_{\mu a}$ (or $g_{\mu\nu}$) and its derivative in accordance with two simple rules.

(i) Adjust the total weight of each term in $\mathfrak{L}$ to -1 by adjoining a factor $|\det L_{\mu a}|^{-w}=|\det g_{\mu\nu}|^{-w/2}$ which transforms as a Lorentz scalar with weight w.

(ii) Replace the ordinary derivative $\partial_\mu\psi=\psi_{,\mu}$ wherever it occurs by the covariant form

$$\psi_{;a}=L^\mu{}_a(\partial_\mu\psi-\tfrac{1}{2}iB_{\mu[bc]}S^{bc}\psi+wL^{\nu b}L_{\nu b,\mu}\psi)\,, \tag{2.8}$$

where S^{bc} denotes the Lorentz spin matrix appropriate to ψ and $B_{\mu[ab]}$, the nonlinear Riemannian connection, denotes the combination

$$\begin{aligned}B_{\mu[ab]}=\tfrac{1}{2}(L^\nu{}_a\partial_\mu L_{\nu b}-L^\nu{}_b\partial_\mu L_{\nu a})-\tfrac{1}{2}(L^\nu{}_a\partial_\nu L_{\mu b}-L^\nu{}_b\partial_\nu L_{\mu a})\\ -\tfrac{1}{2}L_{\mu c}(\partial_\lambda L_\nu{}^c-\partial_\nu L_\lambda{}^c)L^\lambda{}_aL^\nu{}_b\,.\end{aligned} \tag{2.9}$$

The matrix reciprocal to $L_{\mu a}$ is here denoted by $L^{\mu a}$. This notation is consistent with the convention that

[5] A fuller discussion of this point of view is contained in C. J. Isham, Abdus Salam, and J. Strathdee, Ann. Phys. (N. Y.) (to be published). See also B. S. and C. DeWitt, Phys. Rev. 87, 116 (1952); V. I. Ogievetskii and I. V. Polubarinov, Zh. Eksperim. i Teor. Fiz 48, 1625 (1965) [Soviet Phys. JETP 21, 1093 (1965)].

Greek indices are to be raised with $g^{\mu\nu}$ and lowered with $g_{\mu\nu}$ while Latin indices are raised and lowered with the Minkowskian tensor $\eta_{ab}=\eta^{ab}$. Thus we have

$$\begin{aligned} &g^{\mu\nu}g_{\nu\lambda}=\delta_\lambda{}^\mu,\\ &L^{\mu a}L_{\mu b}=\delta_b{}^a,\\ &L^{\mu a}L_{\nu a}=\delta_\nu{}^\mu,\\ &L^{\mu a}g_{\mu\nu}=L_\nu{}^a=L_{\nu b}\eta^{ba}, \text{ etc.} \end{aligned} \tag{2.10}$$

The connection $B_{\mu[ab]}$ is related to the Riemannian connection

$$L_{\mu\nu}{}^\lambda=\tfrac{1}{2}g^{\lambda\rho}(g_{\rho\mu,\nu}+g_{\rho\nu,\mu}-g_{\mu\nu,\rho}) \tag{2.11}$$

by the formula

$$B_{\mu[ab]}=\Gamma_{\mu\nu}{}^\lambda L^\nu{}_a L_{\lambda b}-L^\nu{}_a L_{\nu b,\mu}. \tag{2.12}$$

To the Lagrangian generated by means of the rules (i) and (ii) it is of course necessary to add a purely gravitational term, viz.,

$$\mathfrak{L}_{\text{grav}}=(1/\kappa^2)(-\det g_{\mu\nu})^{1/2}R=(1/\kappa^2)(-\det L_{\mu a})R, \tag{2.13}$$

where R denotes the scalar curvature which can be expressed in terms of either the metric tensor $g_{\mu\nu}$ or the vierbein components $L_{\mu a}$.

It is well known that the ten equations of motion which are obtained by varying $g_{\mu\nu}$ or the (symmetric) $L_{\mu a}$ are not all independent. They satisfy four identities as a consequence of general covariance. In other words, only six of the ten components of $g_{\mu\nu}$ can be determined by these equations. In order to pick out a unique solution it is necessary to supplement the equations of motion with a set of four "coordinate conditions" (which are analogs of the gauge condition of electrodynamics). For example, one could impose the Fock–de Donder conditions

$$\partial_\mu[(-\det g)^{1/2}g^{\mu\nu}]=0 \tag{2.14}$$

on the metric tensor or, alternatively,

$$\partial_\mu[(-\det L)^{1/2}L^{\mu a}]=0 \tag{2.15}$$

on the vierbein components. These classical coordinate conditions have a counterpart in the quantized theory which we shall now discuss.

In setting up the quantized theory one can allow for the fact that the vacuum expectation value of the metric tensor coincides with the flat space form by writing

$$g^{\mu\nu}=\eta^{\mu\nu}+\kappa h^{\mu\nu} \tag{2.16}$$

and treating $h^{\mu\nu}$ as the graviton field. This represents only one of various equivalent schemes. One could take instead of $g^{\mu\nu}$ any second-rank object made out of the metric tensor, its square root $L_{\mu a}$, or its inverse $g_{\mu\nu}$ multiplied into some power of $|\det g^{\mu\nu}|$. These different choices for the basic graviton field would yield different expressions for the Green's functions but presumably identical results for the on-mass-shell S-matrix elements. On the other hand, the perturbation developments of S-matrix elements may converge more rapidly for some choices than for others.

To construct the perturbation series one must separate from L_{grav} the terms which are quadratic in h. These constitute the free Lagrangian L_0. Corresponding to the parametrization (2.16), one finds for L_0 the expression

$$\begin{aligned} \mathfrak{L}_0=\tfrac{1}{4}(&\partial_\mu h^{\lambda\rho}\partial_\mu h^{\lambda\rho}-2\partial_\rho h^{\lambda\mu}\partial_\lambda h^{\rho\mu}\\ &-\partial_\lambda h^{\mu\mu}\partial_\lambda h^{\rho\rho}+2\partial_\mu h^{\mu\lambda}\partial_\lambda h^{\rho\rho}), \end{aligned} \tag{2.17}$$

where contractions relative to the Minkowski metric are tacitly implied, e.g., $h^{\mu\mu}=h^{00}-h^{11}-h^{22}-h^{33}$. The remainder $L_{\text{grav}}-L_0$ is to be treated as an interaction Lagrangian.

The free Lagrangian (2.17) is degenerate (in the sense that it yields an underdetermined set of field equations) owing to the general covariance of the system. In order to define a bare-graviton propagator it is necessary to take account of this general covariance by imposing a set of coordinate conditions. One way to achieve this is by means of a Lagrange multiplier method originated by Fradkin.[6] To L_{grav} one adds the noncovariant term

$$\begin{aligned} -\tfrac{1}{2}\bar{B}_\mu\bar{B}_\mu+\frac{1}{\kappa}\bar{B}_\nu\partial_\mu[(-g)^{1/2}g^{\mu\nu}]&=-\tfrac{1}{2}\bar{B}_\mu\bar{B}_\mu\\ &+\frac{1}{\kappa}\bar{B}_\nu\partial_\mu\left[\frac{\eta^{\mu\nu}+\kappa h^{\mu\nu}}{[-\det(\eta+\kappa h)]^{1/2}}\right], \end{aligned} \tag{2.18}$$

where $\bar{B}_\mu$ denotes the nonlocal expression

$$\bar{B}_\mu(x)=\int d_4y\ \mathfrak{D}_{\mu\nu}(x,y|h)B_\nu(y),$$

with a local field $B_\mu(x)$ to be varied independently of $h^{\mu\nu}$. The kernel $\mathfrak{D}_{\mu\nu}$ is determined by requiring that the equation

$$\partial^2 B_\mu(x)=0 \tag{2.19}$$

emerge as one of the equations of motion. Since this artificial field B_μ is free, one can pick out the space of physical states by means of the condition

$$B_\mu{}^{(+)}(x)|\ \rangle=0 \tag{2.20}$$

and be assured that the S matrix is unitary in this space.

Having determined $\mathfrak{D}$, one can proceed to decouple the artificial field $B(x)$ by means of a field transformation

$$B_\mu\to B_\mu'=\mathfrak{D}_{\mu\nu}B_\nu-(1/\kappa)\eta_{\mu\alpha}\partial_\beta[(-g)^{1/2}g^{\alpha\beta}]. \tag{2.21}$$

Since this transformation is a nonlocal one, it yields a Jacobian factor with a nonlocal structure which must be taken into account in computing S-matrix elements.

[6] E. S. Fradkin and I. V. Tyutin, Phys. Rev. D 2, 2841 (1970); Abdus Salam and J. Strathdee, *ibid.* 2, 2869 (1970).

Thus

$$\left|\det\left(\frac{\delta B_\mu(x)}{\delta B_\nu'(x')}\right)\right| = \exp[-\mathrm{Tr}\ln\mathfrak{D}_{\mu\nu}(x,x'|h)], \quad (2.22)$$

which can be evaluated by perturbation methods. The perturbation development of this functional can be represented graphically by a set of diagrams in which external graviton lines are coupled to closed disjoint loops of a massless vector particle. We shall not go into the rather complicated details[7] here since in fact we shall have no occasion to include the fictitious particle loops in the processes to be considered in Sec. IV. This discussion was intended only to make plausible the effective gravitational action

$$\int dx\left\{\mathfrak{L}_{\mathrm{grav}}+\frac{1}{2\kappa^2}[\partial_\mu((-g)^{1/2}g^{\mu\nu})]^2-i\,\mathrm{Tr}\ln\mathfrak{D}\right\}, \quad (2.23)$$

from which one can separate the bilinear terms which define a nondegenerate free-graviton Lagrangian,

$$\mathfrak{L}_0=\tfrac{1}{4}(\partial_\mu h^{\lambda\rho}\partial_\mu h^{\lambda\rho}-\tfrac{1}{2}\partial_\mu h^{\lambda\lambda}\partial_\mu h^{\rho\rho}). \quad (2.24)$$

The bare-graviton propagator which corresponds to this Lagrangian is simply

$$\langle T(h^{\lambda\mu}(x)h^{\nu\rho}(0))\rangle = \tfrac{1}{2}(\eta^{\lambda\nu}\eta^{\mu\rho}+\eta^{\lambda\rho}\eta^{\mu\nu}-\eta^{\lambda\mu}\eta^{\nu\rho})D(x), \quad (2.25)$$

where D denotes the zero-mass causal propagator $(-4\pi^2x^2)^{-1}$.

Similar arguments can be applied when the graviton is interpolated by the vierbein field $L^{\mu a}$ rather than $g^{\mu\nu}$. One must substitute

$$L^{\mu a}=\eta^{\mu a}+\tfrac{1}{2}\kappa\phi^{\mu a} \quad (2.26)$$

and collect the terms bilinear in h in order to define the bare-graviton propagator. Clearly these bilinear terms are the same as those obtained above, so that (2.25) remains the bare-graviton propagator. The interaction Lagrangian, which involves terms of the third and higher orders, will be different.

III. GRAVITON SUPERPROPAGATOR

The Lagrangian of gravity-modified electrodynamics is given by

$$\mathfrak{L}=\mathfrak{L}_{\mathrm{grav}}+\mathfrak{L}_{\mathrm{matter}}, \quad (3.1)$$

where the purely gravitational part $\mathfrak{L}_{\mathrm{grav}}$ is the usual Einstein one,

$$\mathfrak{L}_{\mathrm{grav}}=\frac{1}{\kappa^2\det L}R(L) = \frac{L^{\mu a}L^{\nu b}}{\kappa^2\det L}(B_{\mu ac}B_{\nu cb}-B_{\nu ac}B_{\mu cb}) +(\text{four-divergence}), \quad (3.2)$$

where the scalar curvature R has been expressed as a function of the vierbein components. The vierbein connection $B_{\mu ab}$ is given by (2.9) as a function of L and its first derivatives. The matter part of (3.1) is given by

$$\mathfrak{L}_{\mathrm{matter}}=(1/\det L)[\tfrac{1}{2}i(\bar\psi\gamma_a\psi_{;\mu}-\bar\psi_{;\mu}\gamma_a\psi)L^{\mu a}-m_0\bar\psi\psi +e_0\bar\psi\gamma_a A_\mu\psi L^{\mu a}-\tfrac{1}{4}L^{\mu a}L^{\nu a}L^{\kappa b}L^{\lambda b}F_{\mu\kappa}F_{\nu\lambda}] +3i\delta^4(0)\ln|\det L|, \quad (3.3)$$

where

$$\psi_{;\mu}=\partial_\mu\psi-\tfrac{1}{4}iB_{\mu ab}\sigma_{ab}\psi, \qquad F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu. \quad (3.4)$$

The general covariance of (3.3) and (3.4) is assured if both the electron and photon fields are assigned the weight $w=0$ [see Eq. (2.6)]. In the above expressions we are using the notational convention that repeated Latin indices are summed in the Minkowskian sense: $A_aB_a=\eta^{ab}A_aB_b=A^aB_a$, etc. The term proportional to $\delta^4(0)$ is due to the presence of a derivative of the electron field in the interaction. It can be justified by a canonical quantization procedure (Salam and Strathdee[6]). There could well be an analogous term in the purely gravitational Lagrangian but we have not been able to derive its form yet.

In order to complete the system of equations it is necessary, as outlined in Sec. II, to add to (3.1) a term which breaks the electromagnetic and gravitational gauge symmetries. Such a term is given by

$$\mathfrak{L}_{\mathrm{gauge}}=-\frac{1}{2\det L}(L^{\mu a}L^{\nu a}A_{\mu;\nu})^2 +\frac{2}{\kappa^2}\left\{\partial_\mu\left[\frac{L^{\mu a}}{(-\det L)^{1/2}}\right]\right\}^2. \quad (3.5)$$

This choice of symmetry-breaking term leads to the following expressions for the bare propagators of the photon and graviton, respectively:

$$\langle 0|T(A_\mu(x)A_\nu(0))|0\rangle=-\eta_{\mu\nu}D(x), \qquad \langle 0|T(\phi^{\mu a}(x)\phi^{\nu b}(0))|0\rangle =\tfrac{1}{2}(\eta^{\mu\nu}\eta^{ab}-\eta^{\mu a}\eta^{\nu b}+\eta^{\mu\nu}\eta^{ab})D(x), \quad (3.6)$$

where $D(x)$ denotes the zero-mass causal propagator $[-4\pi^2(x^2-i0)]^{-1}$. The graviton field $\phi^{\mu a}$ is defined in terms of the vierbein components by

$$L^{\mu a}=\eta^{\mu a}+\tfrac{1}{2}\kappa\phi^{\mu a}. \quad (3.7)$$

With these definitions it is possible to proceed in the usual way to separate from $\mathfrak{L}+\mathfrak{L}_{\mathrm{gauge}}$ the interaction part and to construct a perturbation series. For our purposes, only one part of the interaction need be considered, viz.,

$$\mathfrak{L}_{\mathrm{em}}=e_0\frac{L^{\mu a}}{\det L}\bar\psi\gamma_a\psi A_\mu, \quad (3.8)$$

which will be used in Sec. IV to obtain a contribution of order e_0^2 to the electron and photon propagators.

[7] R. P. Feynman, Acta Phys. Polon. 24, 697 (1963); B. S. DeWitt, Phys. Rev. 162, 1195 (1967); L. D. Faddeev and V. N. Popov, Phys. Letters 25B, 29 (1967); S. Mandelstam, Phys. Rev. 175, 1604 (1968); E. S. Fradkin and I. V. Tyutin, Ref. 6.

There will of course be other contributions of this order due to gravitational couplings of the electron and photon fields. These, which are presumably of higher order in the gravitational constant κ^2, we shall disregard. From (3.8) one obtains for the electron self-energy part

$$(1/i)\Sigma(x)=e_0^2\mathfrak{D}^{\mu a,\nu b}(x)\gamma_a S(x)\gamma_b D_{\mu\nu}(x) \quad (3.9)$$

and, for the photon self-energy part,

$$(1/i)\Pi_{\mu\nu}(x)=-e_0^2\mathfrak{D}^{\mu a,\nu b}(x) \times \mathrm{Tr}[\gamma_a S(x)\gamma_b S(-x)], \quad (3.10)$$

where $S(x)$ and $D_{\mu\nu}(x)$ denote the electron and photon bare propagators, respectively. The graviton superpropagator $\mathfrak{D}^{\mu a,\nu b}(x)$ which appears in these expressions is defined by

$$\mathfrak{D}^{\mu a,\nu b}(x)=\langle 0|\,T\left(\frac{L^{\mu a}(x)}{\det L(x)}\frac{L^{\nu b}(0)}{\det L(0)}\right)|0\rangle, \quad (3.11)$$

where $L^{\mu a}(x)$ is given in terms of the bare-graviton field $\phi^{\mu a}(x)$ by (3.7). The propagator of $\phi^{\mu a}$ is given by (3.6). The construction of the superpropagator from the bare one is a complicated algebraic operation, the main steps of which are sketched in the remainder of this section.

It is convenient to refer the graviton field to a Euclidean basis where the Minkowskian tensor $\eta^{\alpha\beta}$ is replaced by the Kronecker symbol $\delta^{\alpha\beta}$. At the end of the calculation one can transform back to the Minkowskian basis.

The basic step (see Delbourgo and Hunt[8]) in the simplification of (3.11) is to introduce for $(\det L)^{-1}$ the integral representation,

$$\frac{1}{\det L}=\frac{1}{\pi^4}\int d^4m d^4n\,\exp[-(m_\alpha m_\beta+n_\alpha n_\beta)L^{\alpha\beta}], \quad (3.12)$$

where m_α and n_α are integrated over Euclidean four-space. This representation is useful because the chronological pairing of exponentials takes a simple form, viz.,

$$\begin{aligned}&\langle 0|\,T(\exp[-a_{\alpha\beta}L^{\alpha\beta}(x)],\exp[-b_{\gamma\delta}L^{\gamma\delta}(0)])|0\rangle\\&=\exp(-a_{\alpha\alpha}-b_{\alpha\alpha})\exp[a_{\alpha\beta}\tfrac{1}{2}(\delta^{\alpha\gamma}\delta^{\beta\delta}\\&\qquad+\delta^{\alpha\delta}\delta^{\beta\gamma}-\delta^{\alpha\beta}\delta^{\gamma\delta})b_{\gamma\delta}\kappa^2D/4]\\&=\exp(-a_{\alpha\alpha}-b_{\alpha\alpha})\exp[(a_{\alpha\beta}b_{\alpha\beta}-\tfrac{1}{2}a_{\alpha\alpha}b_{\beta\beta})\kappa^2D/4],\end{aligned} \quad (3.13)$$

which can be verified by expanding the exponentials in powers of $\kappa a_{\alpha\beta}\phi^{\alpha\beta}(x)$ and $\kappa b_{\alpha\beta}\phi^{\alpha\beta}(0)$, respectively, and applying (3.6). Differentiation of (3.13) with respect to $a_{\kappa\lambda}$ and $b_{\mu\nu}$ yields the formula

$$\begin{aligned}&\langle 0|\,T(L^{\kappa\lambda}(x)\exp[-a_{\alpha\beta}L^{\alpha\beta}(x)],\\&\qquad L^{\mu\nu}(0)\exp[-b_{\gamma\delta}L^{\gamma\delta}(0)])|0\rangle\\&=\{\tfrac{1}{2}(\delta^{\kappa\mu}\delta^{\lambda\nu}+\delta^{\kappa\nu}\delta^{\lambda\mu}-\delta^{\kappa\lambda}\delta^{\mu\nu})\kappa^2D/4\\&\quad+[\delta^{\kappa\lambda}-(b^{\kappa\lambda}-\tfrac{1}{2}\delta^{\kappa\lambda}b^{\rho\rho})\kappa^2D/4]\\&\quad\times[\delta^{\mu\nu}-(a^{\mu\nu}-\tfrac{1}{2}\delta^{\mu\nu}a^{\rho\rho})\kappa^2D/4]\}\\&\quad\times\langle 0|\,T(\exp[-a_{\alpha\beta}L^{\alpha\beta}(x)],\exp[-b_{\gamma\delta}L^{\gamma\delta}(0)])|0\rangle,\end{aligned}$$

from which one obtains—by substituting $a_{\alpha\beta}=m_\alpha m_\beta+n_\alpha n_\beta$ and $b_{\alpha\beta}=m_\alpha' m_\beta'+n_\alpha' n_\beta'$ and then integrating over all 16 components—an expression for the superpropagator. It is convenient to define a pair of scalar amplitudes $\mathfrak{D}^{(0)}$ and $\mathfrak{D}^{(1)}$ by writing

$$\mathfrak{D}^{\kappa\lambda,\mu\nu}=\delta^{\kappa\lambda}\delta^{\mu\nu}\mathfrak{D}^{(0)}+\tfrac{1}{2}(\delta^{\kappa\mu}\delta^{\lambda\nu}+\delta^{\kappa\nu}\delta^{\lambda\mu}-\delta^{\kappa\lambda}\delta^{\mu\nu})\mathfrak{D}^{(1)}. \quad (3.14)$$

The amplitudes $\mathfrak{D}^{(0)}$ and $\mathfrak{D}^{(1)}$ are now represented formally by 16-dimensional integrals:

$$\begin{aligned}\mathfrak{D}^{(0)}=\frac{1}{\pi^8}\int d^4m d^4n d^4m' d^4n'\{&1+\tfrac{1}{4}(m^2+n^2+m'^2+n'^2)\\&\times\kappa^2D/4+(1/18)(m^2+n^2)(m'^2+n'^2)(\kappa^2D/4)^2\\&+(1/36)[(m\cdot m')^2+(m\cdot n')^2+(n\cdot m')^2\\&+(n\cdot n')^2](\kappa^2D/4)^2\}e^F,\end{aligned} \quad (3.15)$$

$$\begin{aligned}\mathfrak{D}^{(1)}=\frac{1}{\pi^8}\int d^4m d^4n d^4m' d^4n'\{&\kappa^2D/4\\&-(1/36)(m^2+n^2)(m'^2+n'^2)(\kappa^2D/4)^2\\&+\tfrac{1}{9}[(m\cdot m')^2+(m\cdot n')^2+(n\cdot m')^2\\&+(n\cdot n')^2](\kappa^2D/4)^2\}e^F,\end{aligned} \quad (3.16)$$

where

$$\begin{aligned}F=&-(m^2+n^2+m'^2+n'^2)\\&+[(m\cdot m')^2+(m\cdot n')^2+(n\cdot m')^2+(n\cdot n')^2\\&-\tfrac{1}{2}(m^2+n^2)(m'^2+n'^2)]\kappa^2D/4.\end{aligned} \quad (3.17)$$

As they stand, the Riemann integrals (3.15) and (3.16) are divergent. This divergence reflects the fact that one is attempting to sum divergent series [the terms of which can be recovered by expanding the integrands of (3.15) and (3.16) in powers of κ^2D and performing the resulting Gauss-type integrations]. In earlier references[8] the method of Borel summation was used to obtain amplitudes for which the divergent series correspond to asymptotic representations. An equivalent method, which is more convenient in the present case, is to obtain these amplitudes by means of analytic continuation[9] in a set of auxiliary parameters α, β, γ. To this end, consider the integral

$$\begin{aligned}\mathfrak{D}(\alpha,\beta,\gamma)=\frac{1}{\pi^8}\int d^4m d^4n d^4m' d^4n'&\\\times\exp\{-\alpha(m^2+n^2+m'^2+n'^2)&\\+\beta[(m\cdot m')^2+(m\cdot n')^2+(n\cdot m')^2+(n\cdot n')^2]\kappa^2D/4&\\-\tfrac{1}{2}\gamma(m^2+n^2)(m'^2+n'^2)\kappa^2D/4\},&\end{aligned} \quad (3.18)$$

which converges for real and positive values of κ^2D

[8] E. S. Fradkin, Nucl. Phys. 49, 624 (1963); G. V. Efimov, Zh. Eksperim. i Teor. Fiz. 44, 2107 (1963) [Soviet Phys. JETP 17, 1417 (1963)]; R. Delbourgo, Abdus Salam, and J. Strathdee, Phys. Rev. 187, 1999 (1969); R. Delbourgo and A. Hunt, Imperial College, London, Report No. ICTP/69/8 (unpublished).

[9] This technique for defining an integral is a central feature of the method of Gel'fand and Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I. This method was employed particularly by M. K. Volkov, Ann. Phys. (N. Y.) 49, 202 (1968); and also by J. J. Giambiagi and J. Tiomno, Nuovo Cimento Letters 2, 674 (1969); and by Abdus Salam and J. Strathdee, Phys. Rev. D 1, 3296 (1970).

provided that

$$\mathrm{Re}(\beta-\tfrac{1}{2}\gamma)<0 \quad \text{and} \quad \mathrm{Re}\alpha>0. \tag{3.19}$$

The amplitudes $\mathfrak{D}^{(0)}$ and $\mathfrak{D}^{(1)}$ are expressed in terms of $\mathfrak{D}(\alpha,\beta,\gamma)$ and its partial derivatives as follows:

$$\mathfrak{D}^{(0)}=\left[1+\frac{\kappa^2 D}{4}\left(-\frac{1}{4}\frac{\partial}{\partial\alpha}+\frac{1}{36}\frac{\partial}{\partial\beta}-\frac{1}{9}\frac{\partial}{\partial\gamma}\right)\right]\times\mathfrak{D}(\alpha,\beta,\gamma)\big|_{\alpha=\beta=\gamma=1}, \tag{3.20}$$

$$\mathfrak{D}^{(1)}=\frac{\kappa^2 D}{4}\left(1+\frac{1}{9}\frac{\partial}{\partial\beta}+\frac{1}{18}\frac{\partial}{\partial\gamma}\right)\mathfrak{D}(\alpha,\beta,\gamma)\big|_{\alpha=\beta=\gamma=1},$$

where the limit $\alpha=\beta=\gamma=1$ is to be taken in a sense to be specified below.

Many of the integrations in (3.18) are straightforward. Thus, for example, since the exponent in the integrand is a bilinear form in the eight components (m'_α,n'_α), these integrations are Gaussian, yielding

$$\mathfrak{D}(\alpha,\beta,\gamma)=\frac{1}{\pi^4}\int d^4m\,d^4n\{\exp[-\alpha(m^2+n^2)]\}\times\left[\alpha+\frac{\gamma}{2}(m^2+n^2)\frac{\kappa^2 D}{4}\right]^{-2}\left[\alpha^2+\alpha(\gamma-\beta)(m^2+n^2)\frac{\kappa^2 D}{4}+\frac{\gamma}{4}(\gamma-2\beta)(m^2+n^2)^2\left(\frac{\kappa^2 D}{4}\right)^2+\beta^2[m^2n^2-(m\cdot n)^2]\left(\frac{\kappa^2 D}{4}\right)^2\right]^{-1}.$$

Here the integrand depends only on the two combinations

$$\xi=m^2+n^2 \quad \text{and} \quad u^2=4\frac{m^2n^2-(m\cdot n)^2}{(m^2+n^2)^2}$$

and the integral takes the form

$$\mathfrak{D}(\alpha,\beta,\gamma)=\frac{1}{2}\int_0^\infty d\xi\,\xi^3 e^{-\alpha\xi}[\alpha+\gamma\xi(\tfrac{1}{8}\kappa^2 D)]^{-2}\times\int_0^1 du\,u^2\{\alpha^2+2\alpha(\gamma-\beta)\xi(\tfrac{1}{8}\kappa^2 D)+[\gamma(\gamma-2\beta)+\beta^2u^2]\xi^2(\tfrac{1}{8}\kappa^2 D)^2\}^{-1}. \tag{3.21}$$

It is useful to change the integration variables from ξ, u to σ, v, which are defined by

$$v^2=1-u^2, \qquad \sigma=\frac{8\alpha+(\gamma-\beta)\xi\kappa^2 D}{\beta\xi\kappa^2 D}=8\alpha/\beta\xi\kappa^2 D+\sigma_0, \tag{3.22}$$

where

$$\sigma_0=(\gamma-\beta)/\beta.$$

The parameter σ_0, which ultimately takes the value zero, must in the course of the integrations be kept in the half-plane

$$\mathrm{Re}\sigma_0>1.$$

The integral (3.21) now takes the form

$$\mathfrak{D}(\alpha,\beta,\gamma)=\frac{1}{2}\left(\frac{\kappa^2 D}{8}\right)^{-4}\frac{1}{\beta^4}\times\int_{\sigma_0}^\infty\frac{d\sigma}{(1+\sigma)^2}\frac{\exp[-(8\alpha^2/\beta)(\kappa^2 D(\sigma-\sigma_0))^{-1}]}{(\sigma-\sigma_0)}\times\frac{1}{2}\int_0^1 dv^2\frac{(1-v^2)^{1/2}}{\sigma^2-v^2}. \tag{3.23}$$

Of particular interest for the computations of Sec. IV is the Mellin transform of $\mathfrak{D}(\alpha,\beta,\gamma)$ defined by the integral representation

$$\mathfrak{D}(\alpha,\beta,\gamma)=\frac{1}{2\pi i}\int_{C_0} dz\,\Gamma(-z)\mathfrak{D}(\alpha,\beta,\gamma;z)(\kappa^2 D)^z, \tag{3.24}$$

where the contour C_0 lies parallel to the imaginary axis with $-1<\mathrm{Re}z<0$. The amplitude $\mathfrak{D}$ is given by

$$\begin{aligned}\mathfrak{D}(\alpha,\beta,\gamma;z)&=\frac{1}{\Gamma(-z)}\int_0^\infty d(\kappa^2 D)(\kappa^2 D)^{-z-1}\mathfrak{D}(\alpha,\beta,\gamma)\\&=\frac{2^{10}}{\beta^4}\int_0^1 dv^2\int_{\sigma_0}^\infty d\sigma\frac{(1-v^2)^{1/2}}{(1+\sigma)^2(\sigma-\sigma_0)(\sigma^2-v^2)}\frac{1}{\Gamma(-z)}\times\int_0^\infty d(\kappa^2 D)(\kappa^2 D)^{-z-5}\exp\left[-\frac{8}{\kappa^2 D}\frac{\alpha^2/\beta}{\sigma-\sigma_0}\right]\\&=\frac{\Gamma(z+4)}{4\Gamma(-z)}\left(\frac{\beta}{8}\right)^z\left(\frac{1}{\alpha^2}\right)^{z+4}\int_0^1 dv^2(1-v^2)^{1/2}\times\int_{\sigma_0}^\infty d\sigma\frac{(\sigma-\sigma_0)^{z+3}}{(1+\sigma)^2(\sigma^2-v^2)}.\end{aligned} \tag{3.25}$$

To evaluate the σ integral, one can first replace it by a contour integral

$$\int_{\sigma_0}^\infty d\sigma\frac{(\sigma-\sigma_0)^{z-3}}{(1+\sigma)^2(\sigma^2-v^2)}=\frac{1}{2i\sin\pi z}\int_C d\sigma\frac{(\sigma_0-\sigma)^{z+3}}{(1+\sigma)^2(\sigma^2-v^2)},$$

where the contour C is shown in Fig. 1. Since the integrand falls off for large $|\sigma|$ like $|\sigma|^{\mathrm{Re}z-1}$, it is possible to distort the contour so that it encircles only the singularities at -1 and $\pm v$, i.e., into the form C' shown in

Fig. 1—provided Re$z<0$. One finds

$$\frac{1}{\Gamma(-z)}\int_{\sigma_0}^{\infty} d\sigma \frac{(\sigma-\sigma_0)^{z+3}}{(1+\sigma)^2(\sigma^2-v^2)}$$
$$=-\frac{\Gamma(z+1)}{2\pi i}\int_{C'} d\sigma \frac{(\sigma_0-\sigma)^{z+3}}{(1+\sigma)^2(\sigma^2-v^2)}$$
$$=-\Gamma(z+1)\left[\frac{1}{2v}\frac{(\sigma_0-v)^{z+3}}{(1+v)^2}-\frac{1}{2v}\frac{(\sigma_0+v)^{z+3}}{(1-v)^2}\right.$$
$$\left.+\frac{2(\sigma_0+1)^{z+3}}{(1-v^2)^2}-\frac{(z+3)(\sigma_0+1)^{z+2}}{1-v^2}\right],$$

which can be substituted into (3.25) to give

$$\mathfrak{D}(\alpha,\beta,\gamma;z)$$
$$=\tfrac{1}{4}\Gamma(z+1)\Gamma(z+4)(\tfrac{1}{2}\beta)^z(1/\alpha^2)^{z+4}I(z,\sigma_0), \quad (3.26)$$

where

$$I(z,\sigma_0)=-\int_0^1 dv(1-v^2)^{1/2}$$
$$\times\left[\frac{(\sigma_0-v)^{z+3}}{(1+v)^2}-\frac{(\sigma_0+v)^{z+3}}{(1-v)^2}+4(\sigma_0+1)^{z+3}\frac{v}{(1-v^2)^2}\right.$$
$$\left.-2(z+3)(\sigma_0+1)^{z+2}\frac{v}{1-v^2}\right], \quad (3.27)$$

which can be evaluated with the help of a table of integrals. However, for our purposes it will be sufficient to evaluate only the expression

$$I(z)\equiv\tfrac{1}{2}[I(z,+i0)+I(z,-i0)]$$
$$=\int_0^1 dv(1-v^2)^{1/2}\left[\frac{v^{z+3}}{(1+v)^2}\cos\pi z\right.$$
$$\left.+\frac{v^{z+3}}{(1-v)^2}-\frac{4v}{(1-v^2)^2}+\frac{2v(z+3)}{1-v^2}\right]$$
$$=\cos\pi z\, B(\tfrac{3}{2},z+4)F(\tfrac{3}{2},z+4;z+\tfrac{11}{2};-1)$$
$$+B(-\tfrac{1}{2},z+4)F(-\tfrac{1}{2},z+4;z+\tfrac{7}{2};-1)$$
$$-4B(-\tfrac{1}{2},2)F(\tfrac{3}{2},2;\tfrac{3}{2};-1)$$
$$+2(z+3)B(\tfrac{1}{2},2)F(\tfrac{1}{2},2;\tfrac{5}{2};-1). \quad (3.28)$$

This expression corresponds to a particular limiting procedure. Thus, holding α and β fixed at positive real values, the limit $\sigma_0\to\pm i0$ can be interpreted as $\gamma\to\beta\pm i0$. Considered as a function of the complex variable γ, the amplitude $\mathfrak{D}$ has a branch point at $\gamma=2\beta$ with the attached cut running to the left. We are taking the average of the values on the upper and lower sides of this cut in order to have a real amplitude. It may be remarked that the difference of the upper and lower

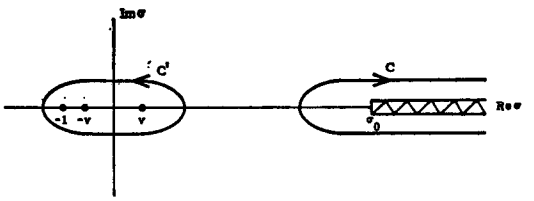


FIG. 1. σ-plane contours for the graviton superpropagator.

limits will vanish at integer values of z and so must correspond to an entire function in momentum space.[10]

In order to construct the Mellin amplitudes corresponding to $\mathfrak{D}^{(0)}$ and $\mathfrak{D}^{(1)}$, it is necessary to differentiate (3.26) with respect to the parameters α, β, and γ, which are then set equal to unity. Using the identity

$$\partial I(z,\sigma_0)/\partial\sigma_0=(z+3)I(z-1,\sigma_0), \quad (3.29)$$

which can be deduced from the integral representation (3.27), one finds

$$\frac{\partial}{\partial\alpha}\mathfrak{D}(z)=-2(z+4)\mathfrak{D}(z),$$
$$\frac{\partial}{\partial\beta}\mathfrak{D}(z)=z\mathfrak{D}(z)-z(z+3)^2\mathfrak{D}(z-1),$$
$$\frac{\partial}{\partial\gamma}\mathfrak{D}(z)=z(z+3)^2\mathfrak{D}(z-1),$$

where

$$\mathfrak{D}(z)=(\tfrac{1}{2})^{z+2}\Gamma(z+1)\Gamma(z+4)I(z).$$

Finally one arrives at the formula

$$\mathfrak{D}^{\mu a,\nu b}(x)=\frac{1}{2\pi i}\int_{C_0} dz\,\Gamma(-z)[\eta^{\mu a}\eta^{\nu b}\mathfrak{D}^{(0)}(z)$$
$$+\tfrac{1}{2}(\eta^{\mu\nu}\eta^{ab}+\eta^{\mu b}\eta^{\nu a}-\eta^{\mu a}\eta^{\nu b})\mathfrak{D}^{(1)}(z)](\kappa^2 D)^z, \quad (3.30)$$

where the amplitudes $\mathfrak{D}^{(0)}(z)$ and $\mathfrak{D}^{(1)}(z)$ are given, respectively, by

$$\mathfrak{D}^{(0)}(z)=\mathfrak{D}(z)-(1/36)z(19z+53)\mathfrak{D}(z-1)$$
$$+(5/36)z(z-1)(z+2)^2\mathfrak{D}(z-2),$$
$$\mathfrak{D}^{(1)}(z)=-\tfrac{1}{9}z(z+8)\mathfrak{D}(z-1) \quad (3.31)$$
$$+(1/18)z(z-1)(z+2)^2\mathfrak{D}(z-2),$$

with $\mathfrak{D}(z)$ given by

$$\mathfrak{D}(z)=(\tfrac{1}{2})^{z+2}\Gamma(z+1)\Gamma(z+4)$$
$$\times[\cos\pi z\, B(\tfrac{3}{2},z+4)F(\tfrac{3}{2},z+4;z+\tfrac{11}{2};-1)$$
$$+B(-\tfrac{1}{2},z+4)F(-\tfrac{1}{2},z+4;z+\tfrac{7}{2};-1)$$
$$-4B(-\tfrac{1}{2},2)F(\tfrac{3}{2},2;\tfrac{3}{2};-1)$$
$$+2(z+3)B(\tfrac{1}{2},2)F(\tfrac{1}{2},2;\tfrac{5}{2};-1)]. \quad (3.32)$$

[10] The possibility of taking the limit $\sigma_0\to 0$ in different ways is a reflection of the arbitrariness in defining T products. See, for example, the discussions by M. K. Volkov (Ref. 9) and B. W Lee and B. Zumino, Nucl. Phys. 13B, 671 (1969).

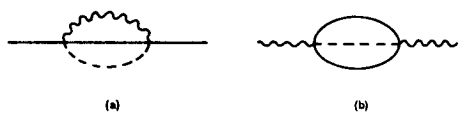


FIG. 2. (a) Electron self-energy Σ; (b) photon self-energy $\Pi_{\mu\nu}$.

Of particular importance for Sec. IV is the behavior near $z=0$, where

$$\mathfrak{D}^{(0)}(z)=1+7.5z+O(z^2),\qquad \mathfrak{D}^{(1)}(z)=-0.5z+O(z^2). \tag{3.33}$$

IV. SECOND-ORDER COMPUTATIONS OF SELF-CHARGE AND SELF-MASS

We now apply the method outlined above to the computation of electromagnetic corrections to the electron and photon masses in the presence of a gravitational field. The graphs to be considered are of order e^2 and are shown in Figs. 2(a) and 2(b). In these figures the electron propagator is represented by the solid line, the photon propagator by the wavy line, and the multigraviton propagator by the dotted line. Corresponding to Fig. 1(a), the configuration space amplitude for electron self-energy is given by [using the approximation of Eq. (3.24)]

$$\begin{aligned}(1/i)\Sigma(x)&=e^2\gamma_\mu S(x)\gamma_\nu D_{\alpha\beta}(x)[\mathfrak{D}^{(0)}(x)\eta_{\mu\alpha}\eta_{\nu\beta}\\&\quad+\tfrac{1}{2}\mathfrak{D}^{(1)}(x)(\eta_{\mu\nu}\eta_{\alpha\beta}+\eta_{\mu\beta}\eta_{\nu\alpha}-\eta_{\mu\alpha}\eta_{\nu\beta})]\\&=\frac{e^2}{2\pi i}\int_{c-i\infty}^{c+i\infty}dz\,\Gamma(-z)[\mathfrak{D}^{(0)}(z)+2\mathfrak{D}^{(1)}(z)]\\&\quad\times\gamma_\mu S(x)\gamma_\mu[\kappa^2D(x)]^z,\end{aligned} \tag{4.1}$$

where $\mathfrak{D}^{(0)}(z)$ and $\mathfrak{D}^{(1)}(z)$ are given by Eq. (3.31). Graphically, the pole $\Gamma(-z)$ at $z=0$ corresponds to the no-graviton exchange contribution, the pole at $z=1$ gives the one-graviton exchange contribution, and so on.

Since, as will be seen in the following, the dominant contribution to Σ comes from the neighborhood of $z=0$, the important quantities are

$$\begin{aligned}&\mathfrak{D}^{(0)}(0)=1,\qquad \mathfrak{D}^{(1)}(0)=0,\\&\frac{d\mathfrak{D}^{(0)}(0)}{dz}=-2\psi(1)+\frac{118\ln2+31}{18}\approx7.5,\\&\frac{d\mathfrak{D}^{(1)}(0)}{dz}=\frac{52\ln2-40}{9}\approx-0.5,\end{aligned} \tag{4.2}$$

where $\psi(a)$ denotes the logarithmic derivative of the Γ function.

Corresponding to Fig. 1(b), the configuration space amplitude for photon self-energy is given by

$$\frac{1}{i}\Pi_{\mu\nu}(x)=\frac{e^2}{2\pi i}\int_{c-i\infty}^{c+i\infty}dz\,\Gamma(-z)[\mathfrak{D}^{(0)}(z)+\tfrac{1}{2}\mathfrak{D}^{(1)}(z)]\times\mathrm{Tr}(\gamma_\mu S(x)\gamma_\nu S(-x))(\kappa^2D)^z, \tag{4.3}$$

where $c<0$.

The momentum space amplitudes $\Sigma(p)$ and $\Pi_{\mu\nu}(k)$ which correspond to (4.1) and (4.3) are given, according to the method of Ref. 9, by the respective contour integrals

$$\Sigma(p)=\frac{1}{2\pi i}\int dz\,\Gamma(-z)[\mathfrak{D}^{(0)}(z)+2\mathfrak{D}^{(1)}(z)]\times\Sigma(p,z), \tag{4.4}$$

$$\Pi_{\mu\nu}(k)=\frac{1}{2\pi i}\int dz\,\Gamma(-z)[\mathfrak{D}^{(0)}(z)+\tfrac{1}{2}\mathfrak{D}^{(1)}(z)]\times\Pi_{\mu\nu}(k,z), \tag{4.5}$$

where $\Sigma(p,z)$ and $\Pi_{\mu\nu}(k,z)$ are defined for sufficiently negative real values of z by the convergent integrals

$$\frac{1}{i}\Sigma(p,z)=\int dx\,e^{ipx}e^2\gamma_\mu S(x)\gamma_\nu D_{\mu\nu}(x)[\kappa^2D(x)]^z, \tag{4.6}$$

$$\frac{1}{i}\Pi_{\mu\nu}(k,z)=\int dx\,e^{ikx}e^2\,\mathrm{Tr}[\gamma_\mu S(x)\gamma_\nu S(-x)]\times[\kappa^2D(x)]^z. \tag{4.7}$$

The contour in (4.4) stands to the left of $z=0$ where, as will be seen below, $\Sigma(p,z)$ has a simple pole. The contour in (4.5) stands to the left of $z=-1$ where $\Pi_{\mu\nu}(k,z)$ has a simple pole.

Into the integrands of (4.6) and (4.7) one can substitute for the propagators as follows:

$$\begin{aligned}S(x)&=(i\gamma_\mu\partial_\mu+m)\frac{mK_1(m\sqrt{(-x^2)})}{4\pi^2\sqrt{(-x^2)}},\\D_{\mu\nu}(x)&=\eta_{\mu\nu}/4\pi^2x^2,\\D(x)^z&=(-1/4\pi^2x^2)^z,\end{aligned} \tag{4.8}$$

and express Σ and $\Pi_{\mu\nu}$ as integrals over the Feynman parameters α, β, and γ. Alternatively, one could perform the integrals directly in configuration space. The results are as follows. Writing

$$\begin{aligned}\Sigma(p,z)&=\not{p}A(p^2,z)+mB(p^2,z),\\\Pi_{\mu\nu}(k,z)&=(k^2\eta_{\mu\nu}-k_\mu k_\nu)C(k^2,z)+\eta_{\mu\nu}D(k^2,z),\end{aligned} \tag{4.9}$$

one finds (the details are given in the Appendix)

$$\begin{aligned}A(p^2,z)&=-\frac{\alpha}{4\pi}\left(\frac{\kappa m}{4\pi}\right)^{2z}\Gamma(2-z)\Gamma(-z)\\&\quad\times{}_2F_1(2-z,-z;3;p^2/m^2),\\B(p^2,z)&=-\frac{\alpha}{\pi}\left(\frac{\kappa m}{4\pi}\right)^{2z}\Gamma(1-z)\Gamma(-z)\\&\quad\times{}_2F_1(1-z,-z;2;p^2/m^2),\end{aligned} \tag{4.10}$$

$$C(k^2,z)=-\tfrac{1}{6}\alpha\pi^{3/2}\left(\frac{\kappa m}{2\pi}\right)^{2z+2}\frac{1}{z}\frac{[\Gamma(2-z)]^2\Gamma(4-z)}{\Gamma(\frac{5}{2}-z)}$$
$$\times{}_3F_2(4-z,2-z,-z;4,\tfrac{5}{2}-z;k^2/4m^2),\quad(4.11)$$
$$D(k^2,z)=-3\alpha\pi^{3/2}\left(\frac{\kappa m}{2\pi}\right)^{2z+2}\frac{1}{z+1}\frac{[\Gamma(1-z)]^2\Gamma(2-z)}{\Gamma(\frac{3}{2}-z)}$$
$$\times{}_3F_2(2-z,1-z,-1-z;3,\tfrac{3}{2}-z;k^2/4m^2),$$

where $\alpha=e^2/4\pi$ denotes the fine-structure constant. These results must be substituted into (4.4) and (4.5) to give the electron and photon self-energies. Since $\Sigma(p,z)$ and $\Pi_{\mu\nu}(k,z)$ contain the factor $(\kappa m)^{2z}$, one can obtain series developments in powers of $(\kappa m)^2$ by shifting the contour to the right.

The leading singularity in the integrand of (4.4) is a dipole at $z=0$ which occurs in both of the scalar amplitudes A and B. The next singularity occurs at $z=1$: a dipole in A and a tripole in B. The remaining singularities at $x=2,3,\ldots$, are all tripoles. The contribution of the leading dipole takes the form

$$A(p^2)\approx-\frac{\alpha}{4\pi}\frac{\partial}{\partial z}\Big[(\mathfrak{D}^{(0)}(z)+2\mathfrak{D}^{(1)}(z))\left(\frac{m}{4\pi}\right)^{2z}\times\frac{\Gamma(2-z)}{\Gamma(z+1)^2}{}_2F_1(2-z,-z;3;p^2/m^2)\Big]_{z=0}$$
$$=\frac{-\alpha}{2\pi}\Big[\ln\left(\frac{4\pi}{\kappa m}\right)+\frac{m^2-p^2}{2p^2}\times\left(1+\frac{m^2+p^2}{p^2}\ln\frac{m^2-p^2}{m^2}\right)+O(1)\Big],$$
$$B(p^2)\approx\frac{\alpha}{\pi}\frac{\partial}{\partial z}\Big[(\mathfrak{D}^{(0)}+2\mathfrak{D}^{(1)})\left(\frac{m}{4\pi}\right)^{2z}\frac{\Gamma(1-z)}{\Gamma(z+1)^2}\times{}_2F_1(1-z,-z;2;p^2/m^2)\Big]_{z=0}\quad(4.12)$$
$$=\frac{2\alpha}{\pi}\Big[\ln\left(\frac{4\pi}{\kappa m}\right)+\frac{m^2-p^2}{2p^2}\times\ln\left(\frac{m^2-p^2}{m^2}\right)+O(1)\Big],$$

which clearly reproduces exactly the amplitudes of ordinary zero-gravity electrodynamics, except that here there remains a dependence on κ in the form of an effective inbuilt cutoff:

$$\Lambda_{\text{cutoff}}\approx 1/\kappa\approx 2\times10^{18}\text{ GeV}.\quad(4.13)$$

In particular,[11] for the mass correction $\Sigma(p)|_{p=m}$ one obtains

$$\frac{\delta m}{m}=\frac{3\alpha}{4\pi}\ln\left(\frac{1}{\kappa m}\right)^2+O((\kappa m)^2\ln(\kappa m)^2).\quad(4.14)$$

Similar remarks apply to the photon self-energy.[11] That is, a dipole at $z=0$ in the transverse part C yields the usual electrodynamic result subject to the effective cutoff (4.13),

$$C(k^2)\approx\tfrac{1}{6}\alpha\pi^{3/2}\frac{\partial}{\partial z}\Big\{[\mathfrak{D}^{(0)}(z)+\tfrac{1}{2}\mathfrak{D}^{(1)}(z)]\left(\frac{m}{2\pi}\right)^{2z+2}\times\frac{\Gamma(2-z)^2\Gamma(4-z)}{\Gamma(z+1)\Gamma(\frac{5}{2}-z)}\times{}_3F_2(4-z,2-z,-z;4,\tfrac{5}{2}-z;k^2/4m^2)\Big\}_{z=0}$$
$$=-\frac{\alpha}{3\pi}\Big[-2\ln\left(\frac{2\pi}{\kappa m}\right)+\left(\frac{2\lambda-1}{2\lambda}\right)\left(\frac{1+\lambda}{\lambda}\right)^{1/2}\times\ln\left(\frac{(1+\lambda)^{1/2}+\sqrt{\lambda}}{(1+\lambda)^{1/2}-\sqrt{\lambda}}\right)+\frac{1}{2\lambda}+O(1)\Big],\quad(4.15)$$

where $\lambda\equiv-k^2/4m^2$. Unfortunately, the longitudinal part $D(k^2)$, which should vanish in a gauge-invariant theory, also has a contribution from simple poles at $z=-1$ and $z=0$:

$$D(k^2)\approx-\alpha\Big\{8\pi\,{}_3F_2(3,2,0;3,\tfrac{5}{2};k^2/4m^2)+\frac{3}{2}\frac{m^2}{\pi}{}_3F_2(2,1;-1;3,\tfrac{3}{2};k^2/4m^2)\Big\}$$
$$=-(\alpha/2\pi)(8\pi^{3/4}/\kappa^2-3m^2+\tfrac{1}{3}k^2),$$

which reflects a gauge noninvariance of our procedure.[12] This will be discussed in Sec. V.

The charge renormalization may be computed from the transverse part $C(k^2)$ of the photon self-energy as

$$\frac{e_R^2}{e_0^2}=Z_3=1-\frac{2\alpha}{3\pi}\ln\frac{2\pi}{\kappa m}+O((\kappa m)^2\ln(\kappa m)^2).$$

The contributions of the tripoles at $z=1,2,\ldots$, to the electron self-energy will give terms like $(\ln\kappa)^2\kappa^{2n}$, $n\geq1$, and can of course be neglected. The important point about the results above is that the ultraviolet singularities of the conventional theory have disappeared via the mechanism of the intrinsic cutoff $1/\kappa$. They still leave their mark, however, as singularities of self-energies in the κ *plane*, i.e., they reappear if the limit $\kappa\to0$ is taken.

[11] Similar formulas have been obtained for a "scalar gravity" model by P. Budini and G. Calucci, Nuovo Cimento **70A**, 419 (1970).

[12] Because of the absence of a $\ln k^2$ term, this non-gauge-invariant part can be removed by a *local* counterterm in the Lagrangian.

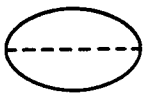


FIG. 3. Electron loop including gravitons.

V. GAUGE INVARIANCE AND EQUIVALENCE THEOREMS

The Lagrangian of gravity-modified electrodynamics given in Sec. III is formally gauge independent. It must follow from this that the scattering amplitudes also are gauge independent provided the computational procedure employs a gauge-invariant regularization. Indeed, if these amplitudes could be expanded in powers of e and κ then they would be gauge independent in each order. However, it is one of the main arguments of this paper that, owing to the appearance of logarithmic singularities at $\kappa=0$, an expansion in powers of κ cannot be made. The question therefore arises as to whether it is necessary to include *all* graphs of order e^l (and these on general grounds must show gauge independence) or whether this collection can be subdivided into sets [possibly of order $e^l(\kappa^2)^m(\ln\kappa)^n$] which are themselves gauge independent. We are at present unable to decide this point and so must confine ourselves to some speculative remarks.

First, the prescription outlined in Sec. III for evaluating graphs with one superpropagator is not a gauge-independent one. This much can be seen from the fact that the vacuum polarization tensor $\Pi_{\mu\nu}(e^2)$ obtained in Sec. IV is not entirely transverse. Although the transverse part of $\pi_{\mu\nu}$ is satisfactory and defines a finite Z_3, the existence of a finite longitudinal component gives a clear indication of the breakdown of gauge symmetry.

If we follow the standard procedure of selecting a gauge-independent set of graphs in electrodynamics, the method is well known and has been formulated by Feynman and Ward. It consists of attaching photon lines in all possible ways to the basic graph of Fig. 3. This then leads to the set of four topologically distinct graphs of Fig. 4. Now, since some of the vertex pairs in these graphs [e.g., pairs 1-3 and 2-3 in Fig. 4(b)] are not connected by superpropagators, we cannot apply the nonpolynomial methods of this paper for computing these contributions. We are in a dilemma now. If we do connect these vertices by graviton superpropagators we can carry through the computation, but the Feynman-Ward procedure for securing gauge invariance would now demand that we include more complicated configurations with three or more superpropagators. For example, graphs depicted in Fig. 5 would have to be added to Fig. 4(b). The resulting graphs will themselves now need new superpropagators, and so on. This "leap-frogging" of the Feynman-Ward procedure for securing gauge invariance, and of joining vertices with superpropagators has only one limit—we must eventually reproduce the entire series $S(e^2)$. This complete series, as stated before, is definitely expected to be gauge invariant. The technical problem—which we have not yet been able to solve—is how to avoid this leap-frogging and to secure gauge invariance without having to sum the entire S-matrix series $S(e^2)$ to second order in e^2 but to all "orders" in κ^2.

Another problem which we have not resolved in this paper is the problem associated with different formulations of gravity theory. We have taken $g^{\mu\nu}$ as the basic field in terms of which tensors like $g_{\mu\nu}$ are to be expressed. One could equally well start with the tensor $g_{\mu\nu}$ as the basic interpolating field and then $g^{\mu\nu}$ would be expressed as a ratio of two polynomials in $g_{\mu\nu}$. If equivalence theorems hold for the type of theory we have been discussing (and these equivalence theorems state that on-mass-shell matrix elements are identical irrespective of which interpolating field we start from, provided that these fields possess the same one-particle asymptotic states and belong to the same locally commutative equivalence class) then the results obtained in this paper should stand, irrespective of which field $g^{\mu\nu}$ or $g_{\mu\nu}$ we choose as basic. The only difference between working with one or the other system of coordinates $g^{\mu\nu}$ or $g_{\mu\nu}$ would be the technical difference of ease in obtaining results more readily in one formulation of the theory relative to the other. The same remark applies to coordinate transformations considered by general relativists by means of which one can incorporate the factor $(\det g)^{-1/4}$ into definitions of matter fields, e.g., $\psi'=(\det g)^{-1/4}\psi$, and thus eliminate this factor from *some* of the terms in the full gravitational-matter Lagrangian. The on-mass-shell results we have obtained should be obtainable, if equivalence theorems hold, by using any set of suitable coordinates, provided that appropriate numbers of terms are added together to secure the equivalence.

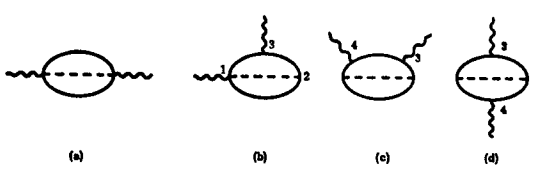


FIG. 4. Gauge-invariant set of graphs to order e^3.

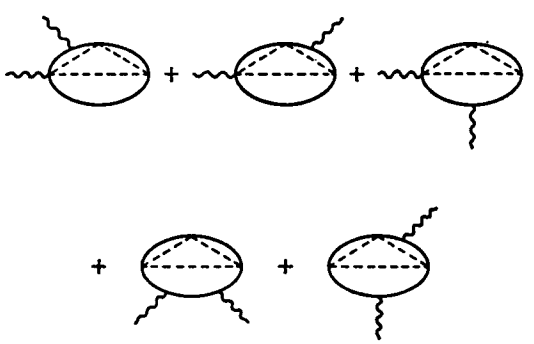


FIG. 5. Superpropagator modifications of Fig. 4.

In this connection it is worth remarking that rational nonpolynomial Lagrangians as a rule give rise to theories which are not strictly localizable (in the sense of Jaffe) when no derivatives appear in the interaction. The derivatives in the gravitational case could change this effect by introducing cancellations. It is, however, also perfectly possible to start with a strictly localizable formulation of gravity theory by parametrizing the vierbein with exponential coordinates, i.e., write, instead of (3.7),

$$L^{\mu a}=[\exp(\tfrac{1}{2}\kappa h)]^{\mu a}, \tag{5.1}$$

where the symmetrical matrix $h^{\mu\nu}$ interpolates the graviton. The exponential parametrization should give a strictly localizable version of gravity *ab initio*. This scheme simplifies some of the computations, e.g.,

$$\langle T \det L(x)^{-1} \det L(0)^{-1}\rangle = \exp[4\kappa^2 D(x)],$$

while making others more difficult. The computations of Sec. IV would be unaffected by this change since the leading terms $O(e^2 \ln\kappa)$ are governed by the normalizing conditions at $z=0$ [cf. Eq. (3.33)],

$$\mathfrak{D}^{(0)}(0)=1, \quad \mathfrak{D}^{(1)}(0)=0,$$

which are unchanged. This equivalence of the two formulations of gravity theory in the lowest order in no way proves their general equivalence. So far as this paper is concerned, this problem is open.

Also open are all problems connected with normal ordering of the gravity-modified Lagrangian. In this paper we have followed a naive ordering prescription although it is known that normal ordering can drastically alter the local-vs-nonlocal properties of field theories, as well as their gauge properties.

Note added in proof. Since this paper was written two important developments have occurred. (1) H. Lehmann and K. Pohlmeyer (private communication) have established analyticity and unitarity for localizable nonpolynomial (and nonderivative) interactions. In addition, they have formulated an asymptotic condition which eliminates the usual distribution-theoretic ambiguities to all orders in the major coupling constant. (2) Assuming that equivalence theorems hold for localizable Lagrangian theories—such as gravity theory in an exponential parametrization Eq. (5.1)—we have succeeded in formulating the computational procedure in such a manner that gauge invariance is preserved to all orders. We have specifically checked that to the order $\alpha \ln k^2 m^2$, there is no change in the numbers given in this paper.

ACKNOWLEDGMENT

It is a pleasure to thank V. F. Weisskopf for a stimulating discussion concerning the significance of the Schwarzschild radius in the present context.

APPENDIX

In this appendix the essential steps involved in the computation of the electron and photon self-energies will be outlined. From Eqs. (5.9) and (4.12) we have

$$A(p^2,z)=\frac{-2e^2}{p^2}\int d^4x\, e^{ip\cdot x}[D(x)]^{z+1}(ip^\mu\partial_\mu)\Delta(x), \tag{A1}$$

$$B(p^2,z)=4e^2\int d^4x\, e^{ip\cdot x}[D(x)]^{z+1}\Delta(x), \tag{A2}$$

where the integrals are taken over the Euclidean region in x-space and are defined initially for $p^2<0$. The angular integrations may be performed immediately using the polar variables

$$dx=4\pi r^2\sin^2\theta dr d\theta,$$
$$p\cdot x=-(-p^2)^{1/2}r\cos\theta,$$
$$(-x^2)^{1/2}=r.$$

We encounter the following integrals:

$$4\pi\int_0^\pi d\theta\sin^2\theta\exp[-i(-p^2)^{1/2}r\cos\theta]=4\pi^2\frac{J_1((-p^2)^{1/2}r)}{(-p^2)^{1/2}r}, \tag{A3}$$

$$4\pi\int_0^\pi d\theta\sin^2\theta\cos\theta\exp[-i(-p^2)^{1/2}r\cos\theta]=4\pi^2\frac{J_2((-p^2)^{1/2}r)}{i(-p^2)^{1/2}r}. \tag{A4}$$

Simplifying the derivative in $A(p^2,z)$ by writing

$$(ip^\mu\partial_\mu)\Delta=2ip^\mu x_\mu\frac{\partial\Delta}{\partial x^2}=2i(-p^2)^{1/2}\frac{\partial\Delta}{\partial r^2}r\cos\theta, \tag{A5}$$

we obtain the expressions

$$A(p^2,z)=16\pi^2e^2\left(\frac{-1}{p^2}\right)\int_0^\infty dr\, r^3J_2((-p^2)^{1/2}r)\times(D)^{z+1}\frac{\partial\Delta(r)}{\partial r^2}, \tag{A6}$$

$$B(p^2,z)=16\pi^2e^2\int_0^\infty dr\, r^3\frac{J_1((-p^2)^{1/2}r)}{(-p^2)^{1/2}}(D)^{z+1}\Delta(r), \tag{A7}$$

into which must be substituted the forms

$$D(r)=(1/4\pi^2)(1/r^2),$$
$$\Delta(r)=(m/4\pi^2)(K_1(mr)/r),$$
$$\frac{\partial\Delta(r)}{\partial r^2}=-\frac{m}{4\pi^2}\frac{K_2(mr)}{2r^2}.$$

The resulting integrals converge for Rez<0 and using the standard relations[13]

$$\int_0^\infty du\, u^{-2z-1}K_1(u)J_1(u(-p^2/m^2)^{1/2})$$
$$=(-p^2/m^2)^{-1/2}4^{-z-1}\Gamma(1-z)\Gamma(-z)$$
$$\times F(1-z,-z;2;p^2/m^2),$$
$$\int_0^\infty du\, u^{-2z-1}K_2(u)J_2(u(-p^2/m^2)^{1/2})$$
$$=(-p^2/2m^2)4^{-z-1}\Gamma(2-z)\Gamma(-z)$$
$$\times F(2-z,-z;3;p^2/m^2),$$

we arrive at Eqs. (4.10).

To evaluate the contribution of the double pole at $z=0$ in Eq. (4.4), it is necessary to evaluate the derivatives of these hypergeometric functions at $z=0$. This may be done by considering the power-series development

$$F(1-z,-z;2;x)=1+\frac{(1-z)(-z)}{2}x$$
$$+\frac{(2-z)(1-z)(1-z)(-z)}{2\times 3}\frac{x^2}{2!}+\cdots,\quad |x|<1$$

from which

$$\left.\frac{\partial F(1-z,-z;2;x)}{\partial z}\right|_{z=0}=-\left(\frac{1\times x}{2}+\frac{2\times 1\times x^2}{2\times 3\times 2!}+\cdots\right)$$
$$=-\sum_{n=1}^{\infty}\frac{x^n}{n(n+1)}. \quad \text{(A8)}$$

This latter sum may readily be shown to be

$$-\{1+[(1-x)/x]\ln(1-x)\}$$

and, substituting this result (with $x=p^2/m^2$), plus a similar one for $F(1-z,-z;3;p^2/m^2)$, into the Sommerfeld-Watson integral, we arrive at the final expression for the electron self-mass given in Eq. (4.12).

The computation of the photon self-energy is somewhat lengthier. The electron calculation was performed directly in configuration space. By way of contrast and also to illustrate the different technique involved, we shall compute the photon self-energy in momentum space using the usual Feynman α-variables. In fact, as a consistency check we have computed the two self-energies both in configuration space and in momentum space.

The α-parameter form of Eq. (4.7), written in terms of the function $C(k^2,z)$ and $D(k^2,z)$, defined in Eq. (4.9) is

$$\Pi_{\mu\nu}(k^2z)=(\eta_{\mu\nu}k^2-k_\mu k_\nu)C(k^2,z)+\eta_{\mu\nu}D(k^2,z),$$

$$C(k^2,z)=\frac{e^2A(z)}{8}\int_0^\infty d\alpha_1\int_0^\infty d\alpha_2$$
$$\times\int_0^\infty d\alpha_3\frac{\alpha_3{}^{3-z}}{\Gamma(2-z)}\frac{\alpha_1\alpha_2}{(\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3)^4}$$
$$\times\exp\left(k^2\frac{\alpha_1\alpha_2\alpha_3}{\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_1\alpha_3}-\alpha_1m^2-\alpha_2m^2\right),$$

$$D(k^2,z)=\frac{e^2A(z)}{16}\int_0^\infty d\alpha_1\int_0^\infty d\alpha_2$$
$$\times\int_0^\infty d\alpha_3\frac{\alpha_3{}^{1-z}}{\Gamma(2-z)}\frac{1}{(\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3)^2}$$
$$\times\left[\frac{\alpha_3}{\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_1\alpha_3}-k^2\frac{\alpha_1\alpha_2\alpha_3{}^2}{(\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_1\alpha_3)^2}+m^2\right]$$
$$\times\exp\left(k^2\frac{\alpha_1\alpha_2\alpha_3}{\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_1\alpha_3}-\alpha_1m^2-\alpha_2m^2\right),$$

where $A(z)=(4\pi)^{-2z}\Gamma(2-z)/\Gamma(z)$. These iterated integrals are defined initially for $k^2<0$ and a certain range of z values. The final answer may be analytically continued to other values.

We shall compute $D(k^2,z)$ first and start by defining new integration variables $x^2=1/\alpha_3$, $y^2=1/\alpha_2$, $z^2=1/\alpha_1$ and then converting to polar coordinates, $x=r\cos\theta$; $y=r\sin\theta\cos\varphi$, $z=r\sin\theta\sin\varphi$. This leads to

$$D(k^2,z)=\frac{e^2A(z)\pi^2}{2\Gamma(2-z)}\int_0^{\pi/2}d\varphi\int_0^{\pi/2}d\theta$$
$$\times\int_0^\infty dr\, r^{2z-1}\sin\varphi\cos\varphi\sin^3\theta(\cos\theta)^{2z-1}$$
$$\times(r^2\sin^4\theta\cos^2\varphi\sin^2\varphi-k^2\sin^4\theta\cos^2\varphi\sin^2\varphi+m^2)$$
$$\times\exp\left(\frac{k^2}{r^2}-\frac{m^2}{r^2\sin^2\theta\cos^2\varphi\sin^2\varphi}\right).$$

The r integration is performed by defining $u=1/r^2$ and then consulting a standard text[13] on Laplace transforms. The result is

$$D(k^2,z)=\frac{e^2A(z)\Gamma(-z-1)}{4\Gamma(2-z)}\int_0^{\pi/2}d\varphi\int_0^{\pi/2}d\theta$$
$$\times\sin\varphi\cos\varphi\sin^3\theta(\cos\theta)^{2z-1}\left(\frac{m^2}{\sin^2\theta\sin^2\varphi\cos^2\varphi}-k^2\right)^z$$
$$\times[\sin^4\theta\sin^2\varphi\cos^2\varphi zk^2-m^2(z+1-\sin^2\theta)],$$

where $k^2<0$ and Rez<1.

[13] W. Erdélyi *et al.*, Bateman Manuscript Project, *Tables of Integral Transforms* (McGraw Hill, New York, 1954), Vols. I and II.

The θ integration may be carried out, using $\sin^2\theta$ as the variable, from the relation[13]

$$\int_0^1 (1-x)^{\mu-1}x^{\nu-1}(x+\alpha)^\lambda dx=\alpha^\lambda\frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}$$
$$\times{}_2F_1(-\lambda,\nu;\mu+\nu,-1/\alpha),\qquad \mathrm{Re}\mu>0,\ \mathrm{Re}\nu>0$$

leading to

$$D(k^2,z)=\frac{e^2A(z)\Gamma(-z-1)\Gamma(z)}{32\Gamma(2-z)}(4m^2)^z\int_0^1 dt(1-t)^{-1/2}$$
$$\times\left\{\frac{z\Gamma(4-z)}{4\Gamma(4)}k^2t^{-z-1}\,{}_2F_1\left(-z,4-z;4;\frac{tk^2}{4m^2}\right)\right.$$
$$+m^2\frac{\Gamma(3-z)}{\Gamma(3)}t^{-z}\,{}_2F_1\left(-z,3-z;3;\frac{tk^2}{4m^2}\right)$$
$$\left.-m^2(z+1)\frac{\Gamma(2-z)}{\Gamma(2)}t^{-z}\,{}_2F_1\left(-z,2-z;2;\frac{tk^2}{4m^2}\right)\right\},$$

where $t=\sin^2 2\varphi$ and $0<\mathrm{Re}z<1$.

Using the standard integral[13]

$$\int_0^1 (1-x)^{\mu-1}x^{\nu-1}\,{}_2F_1(a_1,a_2;b;\alpha x)=\frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}$$
$$\times{}_3F_2(\nu,a_1,a_2;\mu+\nu,\nu,b_1;\alpha),\quad |\alpha|<1$$

we obtain an expression for $D(k^2,z)$ in terms of ${}_3F_2$ functions. This may be usefully converted into one expression in terms of Meyer G-functions, giving

$$D(k^2,z)=\frac{e^2A(z)}{128}\frac{\Gamma(z)}{\Gamma(2-z)}\Gamma(\tfrac{1}{2})\times(-k^2)^{z+1}$$
$$\times\left\{G_{33}^{12}\left(\frac{-k^2}{4m^2}\middle|\begin{matrix}-1, & 1, & -3\\ -z, & -\frac{3}{2}, & -3-z\end{matrix}\right)\right.$$
$$+G_{33}^{12}\left(\frac{-k^2}{4m^2}\middle|\begin{matrix}-1, & 1, & -3\\ -1-z, & -\frac{3}{2}, & -3-z\end{matrix}\right)$$
$$\left.+G_{33}^{12}\left(\frac{-k^2}{4m^2}\middle|\begin{matrix}-1, & 0, & -2\\ -1-z, & -\frac{3}{2}, & -2-z\end{matrix}\right)\right\}.$$

These G-functions may be combined together to give

$$-3zG_{33}^{13}\left(\frac{-k^2}{4m^2}\middle|\begin{matrix}-2, & -1, & 1\\ -z-1, & -z-3, & -\frac{3}{2}\end{matrix}\right).$$

It was for this reason that the Meyer functions were introduced. The final answer, expressed in terms of the more tractable ${}_3F_2$ function, is

$$D(k^2,z)$$
$$=\frac{3e^2}{8\pi^{2+2z}}\frac{(m^2)^{z+1}}{4^z}\frac{\Gamma(\tfrac{1}{2})\Gamma(1-z)\Gamma(1-z)\Gamma(1-z)}{\Gamma(z-1)\Gamma(3)\Gamma(\frac{3}{2}-z)}\frac{\Gamma(z)}{(1+z)}$$
$$\times{}_3F_2(2-z,1-z,-1-z;3,\tfrac{3}{2}-z;k^2/4m^2).$$

The same manipulations for $C(k^2,z)$ yield

$$C(k^2,z)$$
$$=\frac{e^2(m^2)^{z+1}}{16\pi^{2+2z}4^z}\frac{\Gamma(\tfrac{1}{2})\Gamma(1-z)\Gamma(2-z)\Gamma(4-z)}{\Gamma(z-1)\Gamma(4)\Gamma(\frac{5}{2}-z)}\frac{\Gamma(z)}{z}$$
$$\times{}_3F_2(4-z,2-z,-z;4,\tfrac{5}{2}-z;k^2/4m^2),$$

which simplifies slightly to give Eqs. (4.11). The contribution of the double pole at $z=0$ is obtained from the power-series expansion of the ${}_3F_2$ functions as in the electron self-energy calculation.

Volume 40B, number 3 PHYSICS LETTERS 10 July 1972

THE GRAVITATIONAL CORRECTION TO PCAC

R. DELBOURGO and A. SALAM
Physics Department, Imperial College, London S.W. 7 2BZ, U.K.

Received 26 May 1972

A simple-minded perturbation calculation to order G (the Newtonian gravitational constant) of the two-graviton coupling to a pseudoscalar (P) or axial (A) current via a fermion loop gives an anomalous contribution which can be added to the Adler term in the form $\partial_\alpha A_\alpha = 2mP + e^2\epsilon_{\kappa\lambda\mu\nu}F_{\kappa\lambda}F_{\mu\nu}/16\pi^2 + \epsilon_{\kappa\lambda\mu\nu}R_{\kappa\lambda\rho\sigma}R_{\mu\nu\rho\sigma}/768\pi^2$. The anomalous terms can be interpreted as arising from an infinite-mass regulating loop.

The occurrence of anomalous terms in PCAC relations in naive perturbation theory computations involving bilinear currents has prompted us to examine the question for the case when the currents involve derivatives of the fermion fields. In particular we wish to report on the one-fermion loop modifications to such PCAC relations when the stress-tensor current is considered, and especially to consider the basic if academic problem* of the two-graviton decay mode of the π^0 meson, the direct analogue of $\pi^0 \to 2\gamma$. Our result, incorporating the Adler anomaly [1], can be stated in the form

$$\partial_\alpha A_\alpha = 2mP + e^2\epsilon_{\kappa\lambda\mu\nu}F_{\mu\nu}F_{\mu\nu}/16\pi^2 + \\ + \epsilon_{\kappa\lambda\mu\nu}R_{\kappa\lambda\rho\sigma}R_{\mu\nu\rho\sigma}/768\pi^2 \qquad (1)$$

where A is the axial current, P is the pseudoscalar current and R is the Riemann curvature tensor, the gravity analogue of the Maxwell tensor F. All quantities in (1) are in the context of a naive perturbation calculation in lowest order of the relevant couplings (the one-loop graphs), and indeed the anomalous terms can be interpreted as the contribution from an infinite mass fermion regulator in the gravitational as well as the electromagnetic parts. We see in (1) the role played by the two fundamental longrange forces of nature and the parallelism between the Maxwell and Riemann tensors.

In this note we will only mention the salient features of the calculation leaving the considerable details and interpretation of our results to another publication. Because we are dealing with gravity there are two complications * which do not arise with electromagnetism: (a) many more Feynman graphs need to be taken into account when working to any given order in the Newtonian constant $G = 16\pi\kappa^2$, and (b) the resulting integrals are superficially more divergent. Nevertheless, the saving grace which renders subsequent analysis possible and makes the integrals finite is gravitational gauge invariance. In contrast to the two-photon mode of a pseudoscalar or axial current, the two graviton mode brings in three distinct Feynman graphs, as depicted in fig. 1, of which the graviton pole diagram vanishes identically.

With physical gravitons the general form of the matrix element for π^0 decay must, by gravitational gauge invariance, reduce to

$$\epsilon^*_{\mu\nu}(k)\epsilon^*_{\mu'\nu'}(k')\epsilon_{\mu\mu'\lambda\lambda'}k_\lambda k'_{\lambda'}(\eta_{\nu\nu'}kk' - k_{\nu'}k'_\nu)G_P(k\cdot k'). \qquad (2)$$

Likewise the axial (index α) decay must be writable in the form + (3)

$$\epsilon^*_{\mu\nu}(k)\epsilon^*_{\mu'\nu'}(k')\epsilon_{\mu\mu'\lambda\lambda'}k_\lambda k'_{\lambda'}(\eta_{\nu\nu'}k\cdot k' - k_{\nu'}k'_\nu)(k+k')_\alpha G_A(k\cdot k').$$

* This ceases to be an academic problem for the case of strong gravity and the two-f meson decay mode of a (therefore) massive pseudoscalar meson.

* A necessary complication when dealing with spinors is the use of the vierbein formalism.

+ The basic reason for the form (3) is that the 1^+ component of the axial current cannot couple to a pair of identical helicity 2 particles; only the 0^- piece can and this enters in a derivative fashion.

Reprinted with permission from *Phys. Lett.* **40B** (1972) pp. 381–382
© 1972 Elsevier Science Publishers B.V.

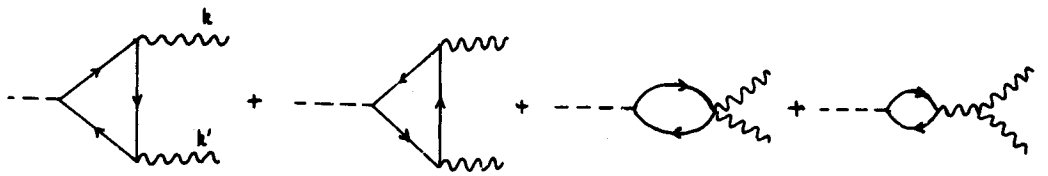


Fig. 1. One-fermion loop contributions to the two-graviton mode.

Expressions (2) and (3) can be directly compared with the matrix elements for the two-photon decay modes [2]

$$\epsilon^*_\mu(k)\,\epsilon^*_{\mu'}(k')\,\epsilon_{\mu\mu'\lambda\lambda'}k_\lambda k'_{\lambda'}F_{\mathrm{P}}(k\cdot k') \quad \text{and}$$

$$\epsilon^*_\mu(k)\,\epsilon^*_{\mu'}(k')\,\epsilon_{\mu\mu'\lambda\lambda'}k_\lambda k'_{\lambda'}(k+k')_\alpha F_{\mathrm{A}}(k\cdot k').$$

Now a naive evaluation of the diagrams of fig. 1 yields*

$$2m\,G_{\mathrm{P}} = \frac{m^2\kappa^2}{8\pi^2}\int_0^1 \frac{\theta(1-x-y)xy\,\mathrm{d}x\,\mathrm{d}y}{(2k\cdot k'xy - m^2)},$$

$$G_{\mathrm{A}} = \frac{-\mathrm{i}\kappa^2}{8\pi^2}\int_0^1 \frac{\theta(1-x-y)x^2y^2\,\mathrm{d}x\,\mathrm{d}y}{(2k\cdot k'xy - m^2)}.$$

We observe an anomaly in that

$$2\mathrm{i}k\cdot k'G_{\mathrm{A}} = 2m\,G_{\mathrm{P}} + \kappa^2/192\pi^2 \neq 2mG_{\mathrm{P}},$$

$$\text{c.f. } 2\mathrm{i}k\cdot k'F_{\mathrm{A}} = 2m\,F_{\mathrm{P}} + e^2/2\pi^2 \neq 2mF_{\mathrm{P}}.$$

By recasting the structure of the effective pseudoscalar Lagrangian in terms of Christoffel symbols { }, we can rewrite the anomalous gravity contribution in terms of the Riemann tensor to this order in G,

$$\epsilon_{\mu\mu'\lambda\lambda'}(\partial_{\nu'}\partial_\lambda g_{\mu\nu}\partial_\nu\partial_{\lambda'}g_{\mu'\nu'} - \partial_\kappa\partial_\lambda g_{\mu\nu}\partial_\kappa\partial_{\lambda'}g_{\mu'\nu'}) =$$

$$= \epsilon_{\mu\lambda\mu'\lambda'}(\partial_\kappa\{\mu\nu,\lambda\}\partial_\kappa\{\mu'\nu,\lambda'\} - \partial_{\nu'}\{\mu\nu,\lambda\}\partial_\nu\{\mu'\nu',\lambda'\}) =$$

$$= \tfrac{1}{2}\epsilon_{\kappa\lambda\mu\nu}R_{\kappa\lambda\rho\sigma}R_{\mu\nu\rho\sigma}$$

finally giving us eq. (1). Furthermore we note that just as the e.m. anomalous term $e^2/2\pi^2 = \lim_{m\to\infty} 2mF_{\mathrm{P}}$ so is the gravitational anomalous term $\kappa^2/192\pi^2 = \lim_{m\to\infty} 2mG_{\mathrm{P}}$, and therefore all anomalous corrections can be identified with contributions from infinite mass regulating loops, which inturn allows for possible re-interpretation of the Ward-Takahashi identity. The attractive possiblity that eq. (1) can be made generally covariant is pure speculation at the present time.

We have greatly benefited from conversations with Dr. C. Isham.

* In deriving the pseudoscalar result, for instance, one encounters a constant nongauge invariant contribution $\eta_{\nu\nu'}\epsilon_{\mu\mu'\lambda\lambda'}k_\lambda k'_{\lambda'}$ which must be eliminated in the same way as one does in the two-photon decay mode of a scalar meson.

References

[1] S. Adler, Phys. Rev. 177 (1969) 2426.
[2] J. Steinberger, Phys. Rev. 76 (1949) 1180; L. Rosenberg, Phys. Rev. 129 (1963) 1786.

Reprinted with permission from *Nuclear Physics* **B76** (1974) 477–482
© 1974 North-Holland Publishing Company

SUPER-GAUGE TRANSFORMATIONS

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and
Imperial College, London, England

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 26 February 1974

Abstract: A systematic method for constructing Wess–Zumino supergauge transformations is exhibited.

In a recent article Wess and Zumino [1] have invented an interesting new symmetry. Generalizing from the dual model super-gauge symmetry [2] these authors succeeded in defining an analogous transformation group in four-dimensional space-time. This invention is quite remarkable in at least two respects: (i) the irreducible representations of this symmetry combine fermions with bosons and (ii) the strictures of O'Raifeartaigh's theorem are circumvented – we seem to have here a relativistic spin-containing symmetry which is consistent with unitarity‡. Moreover, in a simple Lagrangian model involving two scalars and a Majorana spinor, Wess and Zumino found that, in the one-loop approximation, there is only one (logarithmic) divergence [3].

The purpose of this paper is to present a rather simple method which can be used for the construction of at least some of the representations of this symmetry.

We shall confine our considerations to the 14-parameter subalgebra of the Wess–

‡ The group of Wess and Zumino can be looked upon as a sort of quasi U(2, 3): the set of unitary 5 × 5 matrices, g, whose elements g^{β}_{α}, $\alpha, \beta = 1, 2, 3, 4$, and g^{5}_{5} are ordinary complex numbers while g^{5}_{α} and g^{α}_{5} are anti-commuting c-numbers. The subgroup SU(2, 2) × U(1) of the ordinary sort is identified with the product of the 15-parameter conformal group of space-time and a 1-parameter group of γ_5 transformations. The anticommuting parts are identified with supergauge transformations. Looked at in this way an immediate generalization to U(2,4) (or U(2,5)) is suggested. The ordinary subgroup SU(2, 2) × U(2) (or SU(2,2) × U(3)) might then be said to include a strictly internal SU(2) (or SU(3)) symmetry. Contrasted with this marriage of internal symmetries with space-time symmetries, one may also consider a rather trivial generalization of Wess and Zumino's work where each one of their fields is considered as (for example) the adjoint representation of an internal symmetry U(n).

Zumino system which is generated by the Poincare operators $J_{\mu\nu}, P_\mu$ and the Majorana spinor S_α. In addition to the usual commutation rules involving $J_{\mu\nu}$ and P_μ only, we have

$$[S_\alpha, P_\mu] = 0\,,$$

$$[S_\alpha, J_{\mu\nu}] = \tfrac{1}{2}(\sigma_{\mu\nu})_\alpha^\beta S_\beta\,, \tag{1}$$

$$\{S_\alpha, S_\beta\} = (\gamma_\mu C)_{\alpha\beta} P_\mu\,,$$

where C denotes the charge conjugation matrix‡. The last of these rules can be expressed in the alternative version

$$[\bar{\epsilon}_1 S, \bar{S}\epsilon_2] = -\bar{\epsilon}_1 \gamma_\mu \epsilon_2 P_\mu\,, \tag{2}$$

where ϵ_1 and ϵ_2 are two arbitrary Majorana spinors which anticommute with one another and with S. (Notice that $\bar{\epsilon}S = \bar{S}\epsilon$ is a hermitian operator and that $\bar{\epsilon}_1\gamma_\mu\epsilon_2 = -\bar{\epsilon}_2\gamma_\mu\epsilon_1$ is an imaginary 4-vector‡.)

Our basic approach is to work out the group action on the space of left cosets with respect to the subgroup of homogeneous Lorentz transformations. This "space" is essentially eight-dimensional, being parametrized by the 4-vector x_μ and the (anticommuting c-number) Majorana spinor θ_α. A simple way to obtain the group action on this homogeneous space is to define the unitary operators

$$L(x, \theta) = \exp\,[i x_\mu P_\mu]\,\exp\,[i\bar{\theta}^\alpha S_\alpha]\,, \tag{3}$$

and consider what happens to them when any one of the operators representing, respectively, a translation, a homogeneous Lorentz transformation or a super-gauge transformation is applied on the left. One finds,

$$\begin{aligned}
\exp\,[i c_\mu P_\mu]\, L(x, \theta) &= L(x + c, \theta)\,,\\
\exp\,[\tfrac{1}{2} i\omega_{\mu\nu} J_{\mu\nu}]\, L(x, \theta) &= L(\Lambda x, a(\Lambda)\theta)\,\exp\,[\tfrac{1}{2} i\omega_{\mu\nu} J_{\mu\nu}]\,,\\
\exp\,[i\bar{\epsilon} S]\, L(x, \theta) &= L\,[x_\mu - \tfrac{1}{2} i\bar{\epsilon}\gamma_\mu\theta, \theta + \epsilon]\,,
\end{aligned} \tag{4}$$

where $a(\Lambda) = \exp(\tfrac{1}{4} i)\omega_{\mu\nu}\sigma_{\mu\nu}$ denotes the usual spinor representation of the homogeneous Lorentz group‡‡. (Notice that, because of the Majorana constraints on ϵ

‡ Our notational conventions are as follows. The Dirac matrices satisfy $(\tfrac{1}{2})\{\gamma_\mu, \gamma_\nu\} = \eta_{\mu\nu} =$ $=$ diag $(+---)$ and adjoint spinors are defined by $\bar{\psi} = \psi^\dagger\gamma_0$. The matrices γ_0, $\gamma_0\gamma_\mu$, $\gamma_0\sigma_{\mu\nu}$, $\gamma_0 i\gamma_\mu\gamma_5$, $\gamma_0\gamma_5$ are hermitian. The charge conjugate of ψ is defined by $\psi^C = C\bar{\psi}^T$ where $C^T = -C$ and $C^{-1}\gamma_\mu C = -\gamma_\mu^T$. By a Majorana spinor we mean $\psi^C = \psi$. It is useful to remember that the matrices $\gamma_\mu C$ and $\sigma_{\mu\nu}C = (\tfrac{1}{2}i)\,[\gamma_\mu, \gamma_\nu]C$ are symmetric while C, $\gamma_5 C$ and $i\gamma_\mu\gamma_5 C$ are antisymmetric. In particular, it follows that $\bar{\psi}_1\psi_2 = \bar{\psi}_2\psi_1$, $\bar{\psi}_1\gamma_\mu\psi_2 = -\bar{\psi}_2\gamma_\mu\psi_1$, $\bar{\psi}_1\sigma_{\mu\nu}\psi_2 =$ $= -\bar{\psi}_2\sigma_{\mu\nu}\psi_1$, $\bar{\psi}_1 i\gamma_\mu\gamma_5\psi_2 = \bar{\psi}_2 i\gamma_\mu\gamma_5\psi_1$, $\bar{\psi}_1\gamma_5\psi_2 = \bar{\psi}_2\gamma_5\psi_1$ if ψ_1 and ψ_2 are anticommuting Majorana spinors.

‡‡ Space reflections are incorporated by requiring that θ transform according to the rule $\theta \to$ $\to i\gamma_0\theta$. Likewise for dilations, $x \to \lambda x$, $\theta \to \lambda^{\frac{1}{2}}\theta$ and γ_5 transformations, $\theta \to (\cos\alpha + \gamma_5\sin\alpha)\theta$ with real α.

and θ, the displacement in x_μ caused by a super-gauge transformation is real.) Eqs. (4) serve to define the action of the group on the space of the parameters x and θ and indicate how any field defined over this space should transform. Thus, for example, the scalar super-field $\Phi(x, \theta)$ should satisfy

$$\exp[i\bar{\epsilon}S]\,\Phi(x,\theta)\exp[-i\bar{\epsilon}S] = \Phi[x - \tfrac{1}{2}i\bar{\epsilon}\gamma_\mu\theta, \theta + \epsilon]\,. \tag{5}$$

By appending a Lorentz index‡ one could define a vector super-field $\Phi_\mu(x, \theta)$, a spinor super-field $\Psi_\alpha(x, \theta)$, etc.

If we were dealing with an arbitrary group then we should not be very pleased with fields defined on an eight-dimensional space-time. It was this aspect of the old attempts at combining internal symmetries in a non-trivial way with the Poincare group which hindered their development. The truly remarkable and exciting feature of the Wess–Zumino group is that the superfield $\Phi(x, \theta)$ in eight dimensions is exactly equivalent to a 16-component set of ordinary fields in four dimensions. One simply has to expend Φ in powers of θ_α and observe that the series must terminate in the fourth order. This is due, of course, to the fact that the monomials

$$\theta_{\alpha_1}\theta_{\alpha_2}, \ldots, \theta_{\alpha_n}$$

must be completely antisymmetric and therefore vanish for $n > 4$. Therefore we can write

$$\Phi(x,\theta) = \bar{\phi}(x) + \bar{\phi}^{\alpha}(x)\,\theta_\alpha + \tfrac{1}{2}\bar{\phi}^{[\alpha\beta]}(x)\,\theta_\beta\theta_\alpha + \tfrac{1}{6}\bar{\phi}^{[\alpha\beta\gamma]}(x)\theta_\gamma\theta_\beta\theta_\alpha + \tfrac{1}{24}\bar{\phi}^{[\alpha\beta\gamma\delta]}\theta_\delta\theta_\gamma\theta_\beta\theta_\alpha\,. \tag{6}$$

The number of independent real components involved here can be halved by imposing the reality condition

$$\Phi(x,\theta)^* = \Phi(x,\theta)\,, \tag{7}$$

which reads, in terms of the component fields,

$$\begin{aligned}
\phi(x) &= \bar{\phi}(x)\,,\\
\phi_\alpha(x) &= C_{\alpha\alpha'}\bar{\phi}^{\alpha'}(x)\,,\\
\phi_{[\alpha\beta]}(x) &= -C_{\alpha\alpha'}C_{\beta\beta'}\bar{\phi}^{[\alpha'\beta']}(x)\,,\\
\phi_{[\alpha\beta\gamma]}(x) &= -C_{\alpha\alpha'}C_{\beta\beta'}C_{\gamma\gamma'}\bar{\phi}^{[\alpha'\beta'\gamma']}(x)\,,\\
\phi_{[\alpha\beta\gamma\delta]}(x) &= C_{\alpha\alpha'}C_{\beta\beta'}C_{\gamma\gamma'}C_{\delta\delta'}\bar{\phi}^{[\alpha'\beta'\gamma'\delta']}(x)\,,
\end{aligned} \tag{8}$$

‡ The super-gauge transformations induce no Lorentz transformation.

where the barred quantities are defined in the usual way, $\bar{\phi}(x) = \phi(x)^*$, $\bar{\phi}^\alpha(x) = \phi_\beta(x)^*(\gamma_0)^\alpha_\beta$, $\bar{\phi}^{[\alpha\beta]}(x) = \phi_{[\alpha'\beta']}(x)^*(\gamma_0)^\alpha_{\alpha'}(\gamma_0)^\beta_{\beta'}$, etc.

The behaviour of these components under an infinitesimal super-gauge transformation can be extracted from (5). One finds

$$\begin{aligned}
\delta\bar{\phi} &= \bar{\phi}^\alpha \epsilon_\alpha ,\\
\delta\bar{\phi}^\beta &= -\bar{\phi}^{[\alpha\beta]}\epsilon_\alpha - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\beta \partial_\mu \bar{\phi} ,\\
\delta\bar{\phi}^{[\beta\gamma]} &= \bar{\phi}^{[\alpha\beta\gamma]}\epsilon_\alpha + \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\gamma \partial_\mu \bar{\phi}^\beta - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\beta \partial_\mu \bar{\phi}^\gamma ,\\
\delta\bar{\phi}^{[\beta\gamma\delta]} &= -\bar{\phi}^{[\alpha\beta\gamma\delta]}\epsilon_\alpha - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\delta \partial_\mu \bar{\phi}^{[\beta\gamma]} - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\gamma \partial_\mu \bar{\phi}^{[\delta\beta]} - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\beta \partial_\mu \bar{\phi}^{[\gamma\delta]},\\
\delta\bar{\phi}^{[\alpha\beta\gamma\delta]} &= \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\delta \partial_\mu \bar{\phi}^{[\alpha\beta\gamma]} - \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\gamma \partial_\mu \bar{\phi}^{[\delta\alpha\beta]} +\\
&\quad + \tfrac{1}{2}i(\bar{\epsilon}\gamma_\mu)^\beta \partial_\mu \bar{\phi}^{[\gamma\delta\alpha]} - \tfrac{1}{2}(\bar{\epsilon}\gamma_\mu)^\alpha \partial_\mu \bar{\phi}^{[\beta\gamma\delta]} .
\end{aligned} \tag{9}$$

(These rules imply, in particular, that the space-time integral of the component $\phi_{[\alpha\beta\gamma\delta]}$ should be an invariant if surface effects can be neglected.) The representation (9) turns out to be reducible (although not fully reducible). To see this, it is convenient to introduce a new notation for the components. Write

$$\begin{aligned}
\bar{\phi}(x) &= A(x) ,\\
\bar{\phi}^\alpha(x) &= \bar{\psi}^\alpha(x) ,\\
\bar{\phi}^{[\alpha\beta]}(x) &= (C^{-1})^{\alpha\beta}F(x) + (C^{-1}\gamma_5)^{\alpha\beta}G(x) + (C^{-1}i\gamma_\mu\gamma_5)^{\alpha\beta}[a_\mu(x) + \tfrac{1}{2}\partial_\mu B(x)] ,\\
\bar{\phi}^{[\alpha\beta\gamma]}(x) &= \epsilon^{\alpha\beta\gamma\delta}[\lambda_\delta(x) + \tfrac{1}{2}i(\gamma_\mu)^{\delta'}_\delta \partial_\mu \psi_{\delta'}(x)] ,\\
\bar{\phi}^{[\alpha\beta\gamma\delta]}(x) &= \epsilon^{\alpha\beta\gamma\delta}[D(x) - \tfrac{1}{4}\partial^2 A(x)] .
\end{aligned} \tag{10}$$

If the reality conditions (8) are imposed then all boson components are real and the fermion components ψ and λ are Majorana spinors. The axial-vector field a_μ is constrained to be transverse, $\partial_\mu a_\mu = 0$. The transformation rules (9) now take the form‡

$$\begin{aligned}
\delta A &= \bar{\epsilon}\psi ,\\
\delta B &= \bar{\epsilon}\gamma_5\psi - \frac{i}{\partial^2}\bar{\epsilon}\gamma_5\not{\partial}\lambda ,
\end{aligned}$$

‡ Some of the details in these rules are affected by conventions in the definition of C. Our C is defined such that $(\tfrac{1}{2})\epsilon^{\alpha\beta\gamma\delta}C_{\gamma\delta} = -(C^{-1})^{\alpha\beta}$, $(\tfrac{1}{2})\epsilon^{\alpha\beta\gamma\delta}(\gamma_5 C)_{\gamma\delta} = +(C^{-1}\gamma_5)^{\alpha\beta}$ and $(\tfrac{1}{2})\epsilon^{\alpha\beta\gamma\delta}(i\gamma_\mu\gamma_5 C)_{\gamma\delta} = +(C^{-1}i\gamma_\mu\gamma_5)^{\alpha\beta}$.

$$\delta\psi = \tfrac{1}{2}i\partial_\mu A\gamma_\mu\epsilon + \tfrac{1}{2}i\partial_\mu B\gamma_\mu\gamma_5\epsilon + F\epsilon + G\gamma_5\epsilon + a_\mu i\gamma_\mu\gamma_5\epsilon\,,$$

$$\delta F = \tfrac{1}{2}i\bar{\epsilon}\partial\!\!\!/\psi + \tfrac{1}{2}\bar{\epsilon}\lambda\,,$$

$$\delta G = \tfrac{1}{2}i\bar{\epsilon}\gamma_5\partial\!\!\!/\psi + \tfrac{1}{2}\bar{\epsilon}\gamma_5\lambda\,, \tag{11}$$

$$\delta\lambda = D\epsilon + \tfrac{1}{4}i[\partial_\mu a_\nu - \partial_\nu a_\mu]\sigma_{\mu\nu}\gamma_5\epsilon\,,$$

$$\delta a_\mu = \tfrac{1}{2}\bar{\epsilon}\gamma_\mu\gamma_5\lambda - \tfrac{1}{2}i\frac{\partial_\mu\partial_\nu}{\partial^2}\bar{\epsilon}\gamma_\nu\gamma_5\lambda\,,$$

$$\delta D = \tfrac{1}{2}i\bar{\epsilon}\partial\!\!\!/\lambda\,, \tag{12}$$

and it appears that we are dealing with one of those curious representations that does not reduce in the usual way. The eight independent components in the set D, λ and $a_\mu(\partial_\mu a_\mu = 0)$ clearly transform irreducibly. However, unless they are set equal to zero, they involve themselves in the transformations of the other eight components, A, B, ψ, F and G[‡].

The setting to zero of D, λ and a_μ can be viewed as a covariant constraint. In fact one can construct an axial vector which generalizes the well-known Pauli-Lubanski operator,

$$K_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\kappa\rho}P_\nu J_{\kappa\rho} - \tfrac{1}{4}\bar{S}i\gamma_\mu\gamma_5 S\,, \tag{13}$$

whose transverse part is super-gauge invariant. Thus, the antisymmetric tensor

$$K_{\mu\nu} = P_\mu K_\nu - P_\nu K_\mu \tag{14}$$

commutes with both translations and super-gauge transformations. This operator is realized by the following differential expression:

$$K_{\mu\nu}\Phi = (X_{\mu\nu}C)_{\alpha\beta}\frac{\partial}{\partial\theta_\alpha}\frac{\partial\Phi}{\partial\theta_\beta}$$

$$+ i(X_{\mu\nu}\gamma_\rho\theta)_\alpha\frac{\partial}{\partial x_\rho}\frac{\partial\Phi}{\partial\theta_\alpha} - \tfrac{1}{4}\bar{\theta}X_{\mu\nu}\theta\frac{\partial}{\partial x_\rho}\frac{\partial\Phi}{\partial x_\rho}\,, \tag{15}$$

where $X_{\mu\nu} = -(\tfrac{1}{4})(\gamma_\mu\gamma_5\partial_\nu - \gamma_\nu\gamma_5\partial_\mu)$. The equations $K_{\mu\nu}\Phi = 0$ are solved by $D = \lambda = a = 0$.

The combining of representations into products is, at least in some cases, quite easy. One simply multiplies the super-fields. The detailed combinations of compo-

‡ We have adapted our notation here to that of Wess and Zumino, ref. [1]. Thus, our eq. (11) with λ, a_μ and D set equal to zero corresponds to their eq. (8) with $\partial_\mu\alpha = 0$. Similarly, our (12) corresponds to their formulae on p. 48.

nents will be revealed by expanding the result in powers of θ. For example if

$$\Phi_3(x, \theta) = \Phi_1(x, \theta)\, \Phi_2(x, \theta)\,,$$

then, using the components defined by (6),

$$\bar{\phi}^3 = \bar{\phi}_1 \bar{\phi}_2\,,$$

$$\bar{\phi}_3^{\alpha} = \bar{\phi}_1 \bar{\phi}_2^{\alpha} + \bar{\phi}_1^{\alpha} \bar{\phi}_2\,,$$

$$\bar{\phi}_3^{\alpha\beta} = \bar{\phi}_1 \bar{\phi}_2^{\alpha\beta} + \bar{\phi}_1^{\alpha} \bar{\phi}_2^{\beta} - \bar{\phi}_1^{\beta} \bar{\phi}_2^{\alpha} + \bar{\phi}_1^{\alpha\beta} \bar{\phi}_2\,,$$

$$\bar{\phi}_3^{\alpha\beta\gamma} = \bar{\phi}_1 \bar{\phi}_2^{\alpha\beta\gamma} + \sum_{\text{cyc.}} [\bar{\phi}_1^{\alpha} \bar{\phi}_2^{\beta\gamma} + \bar{\phi}_1^{\alpha\beta} \bar{\phi}_2^{\gamma}] + \bar{\phi}_1^{\alpha\beta\gamma} \bar{\phi}_2\,,$$

$$\bar{\phi}_3^{\alpha\beta\gamma\delta} = \bar{\phi}_1 \bar{\phi}_2^{\alpha\beta\gamma\delta} + \sum \bar{\phi}_1^{\alpha} \bar{\phi}_2^{\beta\gamma\delta} + \sum \bar{\phi}_1^{\alpha\beta} \bar{\phi}_2^{\gamma\delta} + \sum \bar{\phi}_1^{\alpha\beta\gamma} \bar{\phi}_2^{\delta} + \bar{\phi}_1^{\alpha\beta\gamma\delta} \bar{\phi}_2\,. \quad (16)$$

In particular, with $\Phi_1 = \Phi_2$ satisfying $K_{\mu\nu}\Phi = 0$, it is a simple matter to show that

$$\bar{\phi}_3^{\alpha\beta\gamma\delta} = \epsilon^{\alpha\beta\gamma\delta} [-\tfrac{1}{4} A\, \partial^2 A + \tfrac{1}{2} \bar{\psi} \partial\!\!\!/ \psi + \tfrac{1}{4} (\partial_\mu B)^2 + F^2 + G^2]$$

$$= \tfrac{1}{2} \epsilon^{\alpha\beta\gamma\delta} [\tfrac{1}{2} (\partial_\mu A)^2 + \tfrac{1}{2} (\partial_\mu B)^2 + \bar{\psi} i \partial\!\!\!/ \psi + 2F^2 + 2G^2 - \tfrac{1}{2} \partial_\mu (A\, \partial_\mu A)]\,. \quad (17)$$

According to the rules (9) this object must transform by a gradient. Its space-time integral is invariant. Wess and Zumino have proposed to use it as a Lagrangian density [1, 3].

The approach discussed in this paper may not prove to be the most serviceable one available but, with the present hazy understanding of this curious and potentially important symmetry, it seems worthwhile to examine every avenue‡.

References

[1] J. Wess and B. Zumino, Nucl. Phys. B70 (1974) 39.

[2] A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86;
P. Ramond, Phys. Rev. D3 (1971) 2415;
Y. Aharonov, A. Casher and L. Susskind, Phys. Letters 35B (1971) 512;
J-L. Gervais and B. Sakita, Nucl. Phys. B34 (1971) 632;
Y. Iwasaki and K. Kikkawa, City College (NY) preprint (1973).

[3] J. Wess and B. Zumino, Phys. Letters 49B (1974) 52.

[4] R. Penrose and M.A.H. MacCullum, Phys. Reports 6 (1973) 241.

‡ From relations (1) and (2), it appears that the Wess–Zumino formalism may have close connections with the Twistor formalism of Penrose (ref. [4]), though the motivations of the two formalisms are apparently completely different.

Reprinted with permission from *Nuclear Physics* **B80** (1974) 499–505
© 1974 North-Holland Publishing Company

UNITARY REPRESENTATIONS OF SUPER-GAUGE SYMMETRIES

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 14 June 1974

Abstract: A method is given for constructing some of the unitary irreducible representations of the Wess–Zumino super-gauge symmetry. Application of this symmetry to the analysis of S-matrix elements is considered. A new super-gauge symmetry which includes isospin is introduced and some of its representations are constructed.

The super-gauge operations invented by Wess and Zumino [1] have the remarkable property of transforming bosons into fermions and *vice versa*. These authors define such transformations on a multiplet of fields which is then built into an invariant Lagrangian. The absence of ghosts (in the perturbative development at least) shows, among other things, that the symmetry is consistent with unitarity. Indeed, it is possible to proceed directly to the construction of unitary representations. This we shall do in the following.

The super-gauge symmetry of Wess and Zumino may be looked upon as the first example of a unitary and relativistic spin-containing theory. Particles with distinct intrinsic spins are here combined into irreducible multiplets. In this letter we present a generalized super-gauge symmetry which contains isospin as well. It is not a relativistic version of Wigner's SU(4), though it resembles it. It appears to be a potent new symmetry, though we do not speculate at present on its usefulness in particle physics.

The full super-gauge symmetry involves dilatations, conformal and γ_5 transformations. In a recent note [2] we considered the more easily manageable subalgebra which is generated by the Poincare operators $J_{\mu\nu}, P_\mu$ and the Majorana spinor S_α for which we adopted the following algebra:

$$[S_\alpha, P_\mu] = 0\,, \quad [S_\alpha, J_{\mu\nu}] = \tfrac{1}{2}(\sigma_{\mu\nu})_\alpha^\beta S_\beta\,, \quad \{S_\alpha, S_\beta\} = -(\gamma_\mu C)_{\alpha\beta} P_\mu\,. \tag{1}$$

Here C denotes the charge conjugation matrix* and the spinor S_α is constrained to be real in the sense,

$$S_\alpha = C_{\alpha\beta}\bar{S}^\beta \,. \tag{2}$$

Under space reflections it transforms according to**

$$S_\alpha \to i(\gamma_0)_\alpha^\beta S_\beta \,. \tag{3}$$

The above-mentioned generalized super-gauge symmetry is obtained by replacing S_α with the isospinor $S_{\alpha i}$ $(i = 1, 2)$ for which a modified anticommutator is postulated.

$$\{S_{\alpha i}, S_{\beta j}\} = \epsilon_{ij}(i\gamma_\mu\gamma_5 C)_{\alpha\beta}P_\mu \,. \tag{4}$$

The space reflection rule (3) is unchanged but the Majorana constraint (2) is replaced by the SU(2) covariant form

$$S_{\alpha i} = i\epsilon_{ij}(\gamma_5 C)_{\alpha\beta}\bar{S}^{\beta j} \,. \tag{5}$$

The construction of unitary representations begins with the observation that super-gauge transformations must leave invariant the manifold of states with fixed 4-momentum since S_α commutes with P_μ. On this manifold the anticommutator $\{S_\alpha, S_\beta\}$ becomes a fixed set of numbers and we see that the operators generate a Clifford algebra. Since this algebra has just sixteen independent members, its one and only finite-dimensional irreducible representation is in terms of 4 X 4 matrices [3]. As we shall see, these matrices are hermitian (in the sense of (2)) when p_μ is timelike. In this case the super-gauge transformations serve to resolve the manifold of states with fixed p_μ into four-dimensional invariant subspaces. If p_μ is light-like, these four-dimensional spaces involve two states of positive norm and two orthogonal states of zero norm.

One can contemplate a complete classification of the unitary irreducible representations according to whether the four-momentum is timelike, lightlike, spacelike or null. In all but the first case one would in general meet infinite-dimensional representations of the Clifford algebra***. Here we shall be dealing exclusively with timelike representations.

In addition to the operators S_α there are, of course, the well-known rotations of Wigner's little group which leave invariant the manifold of states with fixed p_μ. The

* Our notational conventions are as follows. The Dirac matrices satisfy $\frac{1}{2}\{\gamma_\mu, \gamma_\nu\} = \eta_{\eta\nu} = \text{diag}\,(+ - - -)$. Adjoint spinors are defined by $\bar{\psi} = \psi^+\gamma_0$. The matrices γ_0, $\gamma_0\gamma_\mu$, $\gamma_0\sigma_{\mu\nu}$, $\gamma_0 i\gamma_\mu\gamma_5$, $\gamma_0\gamma_5$ are Hermitian. The antisymmetric matrix C defined by $C^{-1}\gamma_\mu C = -\gamma_\mu^{\mathrm{T}}$ is real.

** For consistency with the Majorana constraint the factor i is necessary.

*** Such representations are relevant for group theoretical analysis in the crossed channels. See for example ref. [4].

generators of these rotations, taken together with the Clifford elements, $S_\alpha, S_{[\alpha}S_{\beta]}$, etc., span an algebra which might be called the little algebra. The basic problem is to find the unitary irreducible representations of this algebra. These representations are in one-one correspondence with the unitary irreducible representations of the full group. Similar considerations apply to the generalization (4), (5) where the irreducible representations of the timelike Clifford algebra are sixteen-dimensional.

In the rest frame, $p_0 = M, \boldsymbol{p} = 0$, the little algebra is generated by the angular momentum operators $\boldsymbol{J}$ and the S_α which here obey the rule

$$\{S_\alpha, S_\beta\} = -M(\gamma_0 C)_{\alpha\beta} . \tag{6}$$

In a basis where γ_0 is diagonal and C is real, the Majorana constraint (2) implies $S_4 = -S_1^+, S_3 = S_2^+$ and the anticommutators (6) can be expressed in the suggestive form:

$$\{S_a, S_b\} = 0 , \qquad \{S_a^+, S_b^+\} = 0 , \qquad \{S_a, S_b^+\} = M\delta_a^b , \tag{7}$$

where S_a is a two-component spinor under space rotations and transforms according to

$$S_a \to iS_a \tag{8}$$

under reflections. Viewed as creation and annihilation operators, the spinors S_a and S_a^+ can be used in the familiar way to set up a four-dimensional "Fock space" with positive metric. The procedure is as follows.

Choose a set of $2\mathcal{J}+1$ vectors $|\mathcal{J}, \mathcal{J}_3\rangle, -\mathcal{J} \leqslant \mathcal{J}_3 \leqslant \mathcal{J}$, to represent the states of a particle at rest with mass M and spin $\mathcal{J}$. Let these states constitute the "Clifford vacuum", i.e.

$$S_a^+ |\mathcal{J}, \mathcal{J}_3\rangle = 0 . \tag{9}$$

Define the orthonormal vectors

$$|\mathcal{J}\mathcal{J}_3 n_1 n_2\rangle = M^{-\frac{1}{2}(n_1+n_2)} S_1^{n_1} S_2^{n_2} |\mathcal{J}\mathcal{J}_3\rangle , \tag{10}$$

with the pair (n_1, n_2) taking the values (0,0), (0,1), (1,0) and (1,1). These states span a $4(2\mathcal{J}+1)$-dimensional irreducible representation of the little algebra*. The $(\text{spin})^{\text{parity}}$ content of the multiplet is $(\mathcal{J}-\frac{1}{2})^\eta, \mathcal{J}^{i\eta}, \mathcal{J}^{-i\eta}, (\mathcal{J}+\frac{1}{2})^\eta$, where η takes one of the values $\pm i$ (for integer $\mathcal{J}$) or ± 1 (for half-integer $\mathcal{J}$).

Basis vectors for the representation of the full group are obtained in the usual way by applying Lorentz boosts to the rest frame states (10),

* The three-vector operator $\boldsymbol{\mathcal{J}} = \boldsymbol{J} - (2M)^{-1} S^+\boldsymbol{\sigma}S$ commutes with the Clifford elements and satisfies the commutation rules of SU(2). It coincides in the rest frame with the transverse part of the generalized Pauli–Lubanski vector, K_μ, introduced in ref. [2]. The singlet $(\boldsymbol{\mathcal{J}})^2$ is a Casimir operator. Its eigenvalues $\mathcal{J}(\mathcal{J}+1)$ serve to label the irreducible representations.

$$|p\mathcal{J}\mathcal{J}_3 n_1 n_2\rangle = U(L_p)|\mathcal{J}\mathcal{J}_3 n_1 n_2\rangle\,, \tag{11}$$

where L_p denotes a three-parameter boost which carries the 4-vector $(M,0)$ into p_μ. The behaviour of these states under super-gauge transformations is easily obtained. Apply S_α on both sides of (11) and take it through the operator $U(L_p)$ using the knowledge that it transforms as a Dirac spinor under the Lorentz group. One finds

$$S_\alpha|p\xi\rangle = \sum_{\xi'} |p\xi'\rangle\langle\xi'|S_\alpha(p)|\xi\rangle\,, \tag{12}$$

where ξ denotes the set of labels in (10) and

$$\langle\xi'|S_\alpha(p)|\xi\rangle = a_\alpha^\beta(L_p)\langle\xi'|S_\beta|\xi\rangle\,, \tag{13}$$

where $a_\alpha^\beta(L_p)$ denotes the spinor representation of the Lorentz transformation L_p. The p-dependence of the $4(2\mathcal{J}+1)$-dimensional matrix $S_\alpha(p)$ is thereby given explicitly.

In defining the action of S_α on two-particle states there is one subtle point which must not be overlooked. Suppose we take $S_\alpha = S_\alpha^{(1)} + S_\beta^{(2)}$ where $\{S_\alpha^{(1)}, S_\beta^{(1)}\} = -(\gamma_\mu C)_{\alpha\beta}p_{1\mu}$, and likewise for $S^{(2)}$. Then, in order to have

$$\{S_\alpha^{(1)} + S_\alpha^{(2)}, S_\beta^{(1)} + S_\beta^{(2)}\} = -(\gamma_\mu C)_{\alpha\beta}(p_1 + p_2)_\mu\,, \tag{14}$$

it is essential that $S^{(1)}$ should *anti*commute with $S^{(2)}$. The correct definition of S_α on the product state $|12\rangle$ is therefore,

$$S_\alpha|12\rangle = |1'2\rangle\langle 1'|S_\alpha(p_1)|1\rangle \pm |12'\rangle\langle 2'|S_\alpha(p_2)|2\rangle\,, \tag{15}$$

where the negative sign is used when the state $|1\rangle$ is fermionic.

The symmetry discussed here is a very potent one. To analyse two-body amplitudes, for example, one would resolve the in- and out-states into irreducible representations of the (common) centre-of-mass algebra (14). In general this algebra is smaller than the little algebra of one-particle states, because the presence of relative momentum vectors in the two-particle states forbids the inclusion of J. In general one can supplement (14) only by the discrete transformation corresponding to reflection in the interaction plane. (Forward amplitudes have the additional symmetry against rotations about the common direction of the relative momenta.) Consider, for example, the elastic scattering of two $\mathcal{J} = 0$ multiplets. The sixteen in-states must resolve into four quartets of the algebra (14). Two of these will be even under the discrete reflection and two odd. Likewise for the out-states. It follows that a total of eight amplitudes (four of which vanish in the forward direction) describe the various processes. This count would be further reduced if two or more of the multiplets are identical.

We now consider briefly the application of these ideas to the isospin-containing symmetry characterized by (4) and (5). In this case the Clifford algebra contains

256 independent members, and its fundamental representation is in terms of 16 × 16 matrices. As before, we express the rest-frame anti-commutation rules in the form

$$\{S_{ai}, S_{bj}\} = 0\,, \qquad \{S^{+}_{ai}, S^{+}_{bj}\} = 0\,, \qquad \{S_{ai}, S^{+}_{bj}\} = M\delta^{b}_{a}\delta^{j}_{i}\,, \tag{16}$$

where the four-component object S_{ai} is a spinor under both space and isospace rotations. Since the little algebra now includes I as well as J, we construct its irreducible representations by starting with a multiplet of $(2\mathcal{J}+1)(2\mathcal{I}+1)$ states, $|\mathcal{J}\mathcal{J}_3, \mathcal{I}\mathcal{I}_3\rangle$, and applying the creation operator S_{ai} repeatedly. The resulting states span a space of $16(2\mathcal{J}+1)(2\mathcal{I}+1)$ dimensions with positive metric. These states can then be boosted to span an irreducible unitary representation of the full group. The representations made in this way are seen to be characterized by three numbers, $\mathcal{J}$, $\mathcal{I}$ and M, in addition to parity type*.

In the rest frame the normal parity representation with $\mathcal{I} = \mathcal{J} = 0$ has the $(I, J)^P$ content

$$(0,0)^{+} + (\tfrac{1}{2},\tfrac{1}{2})^{i} + (1,0)^{-} + (0,1)^{-} + (\tfrac{1}{2},\tfrac{1}{2})^{-i} + (0,0)^{+}\,. \tag{17}$$

The content of any other irreducible representation is obtainable from this by vector multiplication with the SU(2) × SU(2) multiplet $(\mathcal{I}, \mathcal{J})$. It will be noticed that the rest frame states (17) can be grouped into multiplets of Wigner's SU(4), *viz.* $1 + 4 + 6 + \bar{4} + 1$. This is because the rest frame Clifford algebra contains the matrices $\frac{1}{2}[S_{ai}, S^{+}_{bj}]$ which obey the SU(4) commutation rules**.

To analyse two-body amplitudes one would resolve the in- and out-states into irreducible representations of the centre-of-mass algebra generated by $S_{\alpha i}$ and I. For example, in the scattering of two $\mathcal{I} = \mathcal{J} = 0$ multiplets the 256 in-states are found to comprise the following representations (denoted $\mathcal{I}^P$) of the centre-of-mass algebra: $1^-, (\frac{1}{2}^{i})^2, (\frac{1}{2}^{-i})^2, (0^+)^2, (0^-)^3$ (where repetitions are indicated by a superscript). One finds a total of twenty-two amplitudes (ten of which vanish in the forward direction) to describe the various processes. At threshold, where the relative momenta vanish, only eight amplitudes survive. Thus, one sees that although the particle multiplets are rather large their scattering appears to be controlled by a relatively manageable number of amplitudes.

* The three-vector $\mathcal{I} = I - (2M)^{-1}S^{+}\tau S$ commutes with the Clifford elements and with $\mathcal{J}$. Its square is a Casimir operator.

** If we had widened the algebra to include three spinors S_{ai} ($i = 1, 2, 3$) and their adjoints, then the algebra of SU(6) would be contained. This programme is not straightforward, however, since the Majorana constraint (5) does not generalize. One may also consider inventing algebras where the anticommutators among the S_{ai}, representing Fermi statistics, are replaced by a more complicated system corresponding to para-statistics (or colour). See note added in proof.

Note added in proof

Generalizations to larger algebraic structures could proceed in at least two distinct ways. The more economical one, which we consider first, involves a set of n Majorana spinors $S_{\alpha k}$ transforming as a vector of O(n). The fundamental anticommutator would take the form:

$$\{S_{\alpha k}, S_{\beta l}\} = -\delta_{kl} (\gamma_\mu C)_{\alpha\beta} P_\mu .$$

The smallest irreducible representation has 2^{2n} components. The rest frame states span a 2^{2n} dimensional Fock space which is obtained by operating repeatedly with the $2n$ "creation operators" S_{ak}, $a = 1, 2$, in the manner explained in the text. These states may be classified according to SU($2n$) which is contained in the little algebra. In the notation of Young tableaux one finds the antisymmetric representations

$$\cdot + \square + \begin{array}{|c|}\hline \\ \hline \\ \hline \end{array} + \begin{array}{|c|c|}\hline \\ \hline \\ \hline \end{array} + \ldots + \left.\begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline \end{array}\right\} 2n .$$

For the case $n = 3$, for example, one finds the SU(6) decomposition

$$2^6 = 1 + 6 + 15 + 20 + \overline{15} + \overline{6} + 1 .$$

Among these states is an O(3) singlet with spin $\frac{3}{2}$. In general, the maximum spin value in the fundamental representation is $\frac{1}{2}n$.

The other (and less economical) scheme is to require the generators $S_{\alpha p}$, $p = 1, 2, \ldots, n$, to transform as an n fold of SU(n). For $n \geqslant 3$ it is necessary to discard the Majorana constraint and treat the generators $\bar{S}^{\alpha p}$ as independent. The fundamental anticommutators would take the form:

$$\{S_{\alpha p}, S_{\beta q}\} = 0, \quad \{\bar{S}^{\alpha p}, S^{\beta q}\} = 0 ,$$

$$\{S_{\alpha p}, \bar{S}^{\beta q}\} = \delta_p^q (\gamma_\mu)_\alpha^\beta P_\mu .$$

In this case the smallest representation has 2^{4n} components since there are now $4n$ independent anticommuting creation operators $S_{\alpha p}$. For the case $n = 3$, the rest frame algebra contains SU(12) which can therefore be used to classify the $2^{12} = 4\,096$ states. One finds all the antisymmetric tensors of SU(12). In terms of dimensions the SU(12) decomposition reads:

$$2^{12} = 1 + 12 + 66 + 220 + 495 + 792 + 924 + \overline{792} + \overline{495} + \overline{220} + \overline{66} + \overline{12} + 1 .$$

In this multiplet one finds spins up to the value $J = 3$.

Although the examples sketched here may not be realistic, they do, at least, show that algebraic generalizations of the Fermi-Bose symmetry are possible. It now becomes important to search out the economical ones.

References

[1] J. Wess and B. Zumino, Nucl. Phys. B70 (1974) 39; Phys. Letters B49 (1974) 52.
[2] Abdus Salam and J. Strathdee, Nucl. Phys. B76 (1974) 477.
[3] J.M. Jauch and F. Rohrlich, Theory of photons and electrons (Addison-Wesley, 1955), appendix A2, p. 425. 1955),
[4] J.F. Boyce, J. Math. Phys. 8 (1967) 675;
A. Sciarrino and M. Toller, J. Math. Phys. 8 (1967) 1252.

Volume 49B, number 5 PHYSICS LETTERS 13 May 1974

ON GOLDSTONE FERMIONS

A. SALAM
International Centre for Theoretical Physics, Trieste, Italy
Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 9 April 1974

Spontaneous violation of super-gauge symmetries is considered. One consequence must be the emergence of a particle with spin 1/2 and mass zero.

In two recent papers [1, 2] Wess and Zumino have generalized the super-gauge symmetry of dual model theory to apply to fields in 4-dimensional space-time. The most unusual feature of super-gauge symmetries is that they combine fermions with bosons in the same multiplet. This means that some of the conserved currents which generate these symmetries must be fermionic. With respect to the Lorentz group, these currents, $J_{\mu\alpha}$, will transform like the product of a vector and a Dirac spinor. Their components are a mixture of spins 1/2 and 3/2.

The unitary irreducible representations [3] of the super-gauge symmetry are characterized by three quantum numbers: the mass M, "spin" J, and parity η. If M is non-vanishing, then the (spin) parity content is $(J-\frac{1}{2})^{\eta}$, $J^{i\eta}$, $J^{-i\eta}$ and $(J+\frac{1}{2})^{\eta}$. If $M=0$, then the helicity content is J, $J+\frac{1}{2}$ and $-J$, $-(J+\frac{1}{2})$.

The purpose of this note is to consider what happens if the vacuum is degenerate, i.e. not a super-gauge singlet. It is natural to expect, on the one hand, a lifting of the mass degeneracy in the previously irreducible multiplets and, on the other, a Goldstone phenomenon. In fact, the immediate implication would be that the system must contain a zero-mass fermion of spin 1/2 because of the following general argument. The action of an infinitesimal super-gauge transformation, ϵ_α, on the Dirac spinor field, ψ_α, is expressed formally by

$$\delta\psi_\alpha(x) = [\psi_\alpha(x), \int \mathrm{d}_3x' \bar{\epsilon}^\beta J_{0\beta}(x')]. \tag{1}$$

Of particular relevance is the vacuum expectation value of this equation, which can be expressed in the form

$$\partial_\mu \langle T^* \bar{\epsilon}^\beta J_{\mu\beta}(x)\, \psi_\alpha(0)\rangle = -\delta_4(x)\langle \delta\psi_\alpha(0)\rangle. \tag{2}$$

This equation is equivalent to (1) if the vacuum is translation invariant and if the current is local and conserved. Now, if the vacuum is not super-gauge invariant, then $\langle\delta\psi_\alpha\rangle$ will not vanish; instead it will have the form

$$\langle\delta\psi_\alpha\rangle = \langle F\rangle \epsilon_\alpha = \langle F\rangle C_{\alpha\beta}\bar{\epsilon}^\beta, \tag{3}$$

where $C_{\alpha\beta}$ denotes the charge conjugation matrix [4]. The non-vanishing coefficient $\langle F\rangle$ which appears here is to be interpreted as the vacuum expectation value of some scalar field to which ψ is related by super-gauge transformations. Thus we have the Ward identity,

$$\partial_\mu \langle T^* J_{\mu\alpha}(x)\, \psi_\beta(0)\rangle = C_{\alpha\beta}\langle F\rangle \delta_4(x). \tag{4}$$

Define the momentum-space transform of this amplitude and resolve it into invariant components,

$$\begin{aligned}\int \mathrm{d}x \exp(ikx)\langle T^* J_{\mu\alpha}(x)\,\psi_\beta(0)\rangle &= M_{\mu\alpha\beta}(k)\\ &= M_1(k^2)k_\mu C_{\alpha\beta} + (M_2(k^2)\eta_{\mu\nu} + M_3(k^2)k_\mu k_\nu)(\gamma_\nu C)_{\alpha\beta}\\ &\quad + M_4(k^2)\,k_\nu(\sigma_{\mu\nu}C)_{\alpha\beta}.\end{aligned} \tag{5}$$

It follows from (4) that the invariant components are constrained by the equations

$$\begin{aligned} k^2 M_1(k^2) &= \mathrm{i}\langle F\rangle\\ M_2(k^2) + k^2 M_3(k^2) &= 0.\end{aligned} \tag{6}$$

Reprinted with permission from *Phys. Lett.* 49B (1974) pp. 465–467
© 1974 Elsevier Science Publishers B.V.

Table 1

	$\langle A\rangle$	$\langle B\rangle$	$\langle F\rangle$	$\langle G\rangle$	m_A^2	m_B^2	m_ψ
$I_\pm$	$-m/g$	$\pm(m/g)\sqrt{g\lambda/m^2-1}$	0	0	$g\lambda-m^2$	$g\lambda-m^2$	$\pm\sqrt{g\lambda-m^2}$
II	$-m/g$	0	$\frac{m^2}{4g}-\frac{\lambda}{4}$	0	$\frac{1}{2}(g\lambda-m^2)$	$-\frac{1}{2}(g\lambda-m^2)$	0
$III_\pm$	$-(m/g)\,[1\pm\sqrt{1-g\lambda/m^2}]$	0	0	0	$m^2-g\lambda$	$m^2-g\lambda$	$\pm\sqrt{m^2-g\lambda}$

The second of these equations serves only to eliminate M_2 from the decomposition (5). The first equation gives the explicit form of $M_1(k^2)$, viz. a simple zero-mass pole with residue $i\langle F\rangle$. This indicates that the intermediate states which contribute to $M_{\mu\alpha\beta}$ must include a massless particle of spin $\frac{1}{2}$: a Goldstone fermion.

The notion of a Goldstone fermion is highly attractive since it might provide a fundamental theory of the neutrino. We have tested the idea in the context of a simple Lagrangian model but have, however, come upon a dilemma, at least so far as the tree approximation goes. The model we consider is that of Wess and Zumino, which employs a multiplet of eight real components: two scalars, A, F, two pseudoscalars, B, G, and a Majorana spinor, ψ. They transform according to

$$\begin{aligned}
&\delta A=\bar\epsilon\psi, \qquad \delta B=\bar\epsilon\gamma_5\psi,\\
&\delta\psi=(F+G\gamma_5)\epsilon-\tfrac{1}{2}i\partial\!\!\!/(A+B\gamma_5)\epsilon,\\
&\delta F=-\tfrac{1}{2}i\bar\epsilon\partial\!\!\!/\psi, \qquad \delta G=-\tfrac{1}{2}i\bar\epsilon\gamma_5\partial\!\!\!/\psi.
\end{aligned}\tag{7}$$

The Lagrangian (which is invariant up to a 4-divergence) is given by

$$L=L_0+L_m+L_g+L_\lambda,\tag{8}$$

where

$$\begin{aligned}
L_0&=\tfrac{1}{2}(\partial_\mu A)^2+\tfrac{1}{2}(\partial_\mu B)^2+\bar\psi i\partial\!\!\!/\psi+2(F^2+G^2)\\
L_m&=2m(AF+BG-\tfrac{1}{2}\bar\psi\psi)\\
L_g&=g[(A^2-B^2)F+2ABG-\bar\psi(A-B\gamma_5)\psi]\\
L_\lambda&=\lambda F.
\end{aligned}\tag{9}$$

The conserved Noether current is given by (suppressing the Dirac index):

$$\begin{aligned}
J_\mu&=\bar\phi(A-B\gamma_5)\gamma_\mu\psi+im(A+B\gamma_5)\gamma_\mu\psi\\
&\quad+\tfrac{1}{2}ig(A+B\gamma_5)^2\gamma_\mu\psi-\tfrac{1}{2}\lambda\gamma_\mu\gamma_5\psi.
\end{aligned}\tag{10}$$

The algebraic variables F and G are redundant. They can be eliminated by using the corresponding Euler-Lagrange equations.‡

$$\begin{aligned}
0&=4F+2mA+g(A^2-B^2)+\lambda\\
0&=4G+2mB+2gAB.
\end{aligned}\tag{11}$$

When this is done the Lagrangian assumes the form

$$\begin{aligned}
L'&=\tfrac{1}{2}(\partial_\mu A)^2+\tfrac{1}{2}(\partial_\mu B)^2\\
&\quad+\bar\psi(i\partial\!\!\!/-m)\psi-g\bar\psi(A-B\gamma_5)\psi-V(A,B),
\end{aligned}\tag{12}$$

where the potential V is given by

$$\begin{aligned}
V&=\tfrac{1}{2}m^2(A+\lambda/2m)^2+\tfrac{1}{2}m^2B^2\\
&\quad+\tfrac{1}{4}g\lambda(A^2-B^2)+\tfrac{1}{2}mgA(A^2+B^2)\\
&\quad+\tfrac{1}{8}g^2(A^2+B^2)^2.
\end{aligned}\tag{13}$$

One can now proceed in the standard fashion to find the extrema of V. The ground states selected in this way must then be tested for stability by computing the particle masses.

There are three solutions. We list in table 1 the values of $\langle A\rangle$, $\langle B\rangle$, $\langle F\rangle$ and $\langle G\rangle$ as well as m_A^2, m_B^2 and m_ψ.

‡ These equations may be used to eliminate F and G from the transformation laws (7) which then become *non-linear*. In other words, the Lagrangian (12) is invariant (up to a 4-divergence) under a non-linear realization of the super-gauge symmetry.

The only stable solutions[‡] are $I_{\pm}$ (for $m^2 < g\lambda$) and $III_{\pm}$ (for $m^2 > g\lambda$). Both of these are symmetry preserving, i.e. $\langle F \rangle = 0$. The Goldstone solution II is unstable since either A or B is a tachyon. We therefore conclude that the Lagrangian (8) cannot support a vacuum asymmetry, at least in the tree approximation here considered.

There are two possible escapes from this dilemma. One, we may need to go beyond the tree approximation [5] [‼]; or alternatively, that we have chosen the wrong super-gauge multiplet (and the wrong Lagrangian) to implement the notion of Goldstone fermions. The particle may in fact belong not to the multiplet with $J = 0$, but to either of the multiplets with $J = \frac{1}{2}$ and $J = 1$.

Note added. We have been informed by Professor B. Zumino that the problem of emergence of Goldstone germions has also been considered by himself and Professor J. Iliopoulos in a forthcoming CERN preprint.

[‡] The two solutions denoted I_+ and I_- are equivalent. One can be brought into the other by a field redefinition. Likewise for III_+ and III_-. The solution I does not violate parity but rather imposes a redefinition of the parity operator. For the special value of λ given by $\lambda = m^2/g$, note, however, the existence of the degenerate solution $\langle A \rangle \neq 0$, but $\langle B \rangle = \langle F \rangle = \langle G \rangle = m_A = m_B = m_\psi = 0$.

[‼] It is well known that in order to generate Goldstone solutions *in a tree approximation*, one must arrange the sequence of signs in the potential in a favourable manner. In the case considered in the text, the super-gauge symmetry has already fixed all the signs for us.

References

[1] J. Wess and B. Zumino, CERN preprint TH.1753 (1973), to appear in Nucl. Phys.
[2] J. Wess and B. Zumino, CERN preprint TH.1794 (1973), to appear in Nucl. Phys.
[3] A. Salam and J. Strathdee, ICTP, Trieste, preprint IC/74/16
[4] For the notational conventions used here see: A. Salam and J. Strathdee, ICTP Triests, preprint IC/74/11.
[5] S. Coleman and E. Weinberg, Phys. Rev. D7 (1973) 1888.

Volume 51B, number 4 PHYSICS LETTERS 19 August 1974

SUPER-SYMMETRY AND NON-ABELIAN GAUGES

A. SALAM
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 6 June 1974

We show that the conventional Yang-Mills Lagrangian with the Yang-Mills field interacting with Majorana fermions belonging to the adjoint representation of an internal symmetry group like SU(n) is super-gauge invariant.

The feasibility of combining fermions with bosons in irreducible multiplets of a new kind of global symmetry (super-gauge symmetry) has been established by Wess and Zumino*. These authors have given two examples [3, 4] of renormalizable Lagrangians which possess this symmetry. The question naturally arises as to whether and to what extent the new symmetry can be compatible with internal symmetries. Internal symmetries of the global type are easily incorporated. More interesting and more difficult is the question of compatibility with internal symmetries of the *local* type. What is clearly implied in such cases is that the vector mesons of the gauge system must be associated with *new gauge particles which are fermions*. In the absence of spontaneous symmetry breaking, the gauge fermions would have zero mass, like the gauge vector mesons.

Wess and Zumino have constructed a super-gauge invariant extension of quantum electrodynamics, i.e. of an Abelian local symmetry [4]. The gauge multiplet consists of a photon together with a neutral (Majorana) zero-mass spinor, while the matter multiplet consists of an electron and two charged spin-zero particles. The resulting model is renormalizable.

In this paper we construct a super-symmetric Lagrangian which at the same time exhibits a *non-Abelian* gauge symmetry. In particular, we find the somewhat remarkable result that the Lagrangian for a conventional Yang-Mills system of fields interacting with a multiplet of Majorana fermions is superinvariant, provided that the Fermi fields (like the Yang-Mills fields themselves) belong to the adjoint representation of the internal symmetry group (e.g. SU(n)).

To construct the Lagrangian, we use the method of super-fields [5]. A typical super-field, $\Phi(x,\theta)$, defined over the space of anticommuting Majorana spinors θ_α, is equivalent to a set of sixteen ordinary fields (half of them fermionic and the other half bosonic) defined over space-time. These ordinary fields are simply the coefficients in the expansion of $\Phi(x,\theta)$ in powers of θ (which expansion must terminate in the fourth order because of the anticommutativity of the variable θ). Differentiation with respect to θ is easily defined and enjoys the usual properties except that attention must be paid to the order of factors. A particularly useful operator is the "covariant derivative" D_α defined by

$$D_\alpha \Phi(x,\theta) = \frac{\partial \Phi}{\partial \bar{\theta}^\alpha} - \frac{i}{2}(\gamma_\mu \theta)_\alpha \frac{\partial \Phi}{\partial x_\mu} . \qquad (1)$$

This derivative is a Dirac spinor under Lorentz trans-

* These authors [1], following the usage in dual model theory where the concept originated [2], designate this Fermi-Bose symmetry by the expression "super-gauge". Since the word "gauge" has come to be associated more commonly with "gauges of the second kind" or local symmetries, it is confusing to use super-gauge to describe what is indeed a global symmetry of fermions and bosons. We suggest therefore that the expression "super-symmetry" might be more appropriate for the global concept and reserve the word "gauge" for local symmetries.

Reprinted with permission from *Phys. Lett.* 51B (1974) pp. 353–355
© 1974 Elsevier Science Publishers B.V.

formations and an invariant under super-transformations,

$$\delta\Phi = -\bar{\epsilon}^{\alpha}\left(\frac{\partial\Phi}{\partial\bar{\theta}^{\alpha}} + \frac{\mathrm{i}}{2}(\gamma_{\mu}\theta)_{\alpha}\frac{\partial\Phi}{\partial x_{\mu}}\right). \tag{2}$$

The fundamental property of the operators D_{α} is that they generate a 16-dimensional Clifford algebra,

$$\{D_{\alpha}, D_{\beta}\} = -(\gamma_{\mu}C)_{\alpha\beta}\mathrm{i}\frac{\partial}{\partial x_{\mu}}. \tag{3}$$

This means, for example, that any product of five or more covariant derivatives can always be reduced*.

A special kind of super-field can be expressed in terms of only four independent components, viz.

$$\Phi_{\pm}(x,\theta) = \exp\left[\mp\tfrac{1}{4}\bar{\theta}\gamma_{\mu}\gamma_{5}\theta\frac{\partial}{\partial x_{\mu}}\right]$$
$$\times\,(A_{\pm}(x) + \bar{\theta}\psi_{\pm}(x) + \tfrac{1}{4}\bar{\theta}(1\pm\mathrm{i}\gamma_{5})\theta F_{\pm}(x)), \tag{4}$$

where $A_{\pm}$ and $F_{\pm}$ are complex, and $\psi_{\pm}$ are chiral projections ($\mathrm{i}\gamma_{5}\psi_{\pm} = \pm\psi_{\pm}$). Clearly, space reflections must carry Φ_{+} into Φ_{-} and vice-versa. These super-fields, which are *irreducible* under the combined super-symmetry and (proper) Poincaré groups, identically satisfy

$$D_{R,L\alpha}\Phi_{\pm} \equiv \tfrac{1}{2}(1\mp\mathrm{i}\gamma_{5})^{\beta}_{\alpha}D_{\beta}\Phi_{\pm} = 0, \tag{5}$$

which can be verified directly using (1) and (4). A fundamental property of the irreducible fields is that the ordinary product of two "left-handed" fields is again left-handed,

$$\Phi'_{+}(x,\theta)\Phi''_{+}(x,\theta) = \Phi'''_{+}(x,\theta), \tag{6}$$

and similarly for the right-handed fields. The product of right with left fields, however, is a general 16-component super-field.

Now consider the action of a local symmetry transformation on a set of super-fields $\Phi(x,\theta)$ which belong to a representation of an internal symmetry group. It is natural to require that the parameters of such a transformation should themselves comprise a super-field,

$$\Phi(x,\theta) \to \Omega(x,\theta)\Phi(x,\theta).$$

The group property

$$\Omega_{1}(x,\theta)\Omega_{2}(x,\theta) = \Omega_{3}(x,\theta)$$

requires that the elements of Ω be super-fields either of the general 16-component type or else of one of the types Ω_{+} or Ω_{-}. In this note we consider the more restricted type of gauge symmetry

$$\Phi_{\pm} \to \exp[\mathrm{i}\Lambda_{\pm}]\Phi_{\pm} \tag{7}$$

where Λ_{+} and Λ_{-} are matrices related by

$$\Lambda_{+}^{\dagger} = \Lambda_{-}. \tag{8}$$

It may be instructive to see the rule (7) expressed in component form with

$$\delta\Lambda_{+} = \exp[-\tfrac{1}{4}\bar{\theta}\not{\phi}\gamma_{5}\theta](\delta\omega_{+} + \bar{\theta}\delta\chi_{+} + \tfrac{1}{4}\bar{\theta}(1+\mathrm{i}\gamma_{5})\theta\delta\phi_{+}). \tag{9}$$

One finds

$$\mathrm{i}^{-1}\delta A_{+} = \delta\omega_{+}A_{+}, \qquad \mathrm{i}^{-1}\delta\psi_{+} = \delta\omega_{+}\psi_{+} + \delta\chi_{+}A_{+},$$
$$\mathrm{i}^{-1}\delta F_{+} = \delta\omega_{+}F_{+} - \delta\bar{\chi}_{-}\psi_{+} + \delta\phi_{+}A_{+}, \tag{10}$$

with similar formulae for the right-handed components. The purely internal symmetry transformations of the fields $A_{\pm}$, $\psi_{\pm}$ and $F_{\pm}$, parametrized by $\delta\omega_{\pm}$, are here seen to be supplemented by parameters $\delta\chi_{\pm}$ and $\delta\phi_{\pm}$ to achieve compatibility with super-symmetry.

The gauge fields are to be contained in a hermitian matrix super-field Ψ which transforms according to

$$\mathrm{e}^{\Psi} \to \mathrm{e}^{\mathrm{i}\Lambda_{-}}\mathrm{e}^{\Psi}\mathrm{e}^{-\mathrm{i}\Lambda_{+}}. \tag{11}$$

The Ψ fields must not be of the Ψ_{+} and Ψ_{-} variety but are general super-fields. The main problem is to construct a gauge-invariant Lagrangian for them. The interaction of Ψ with the matter fields is controlled by the symmetry requirements. The coupling is contained in the invariant expression ‡

$$\mathrm{Tr}\,(\bar{D}D)^{2}(\Phi_{+}^{\dagger}\mathrm{e}^{\Psi}\Phi_{+} + \Phi_{-}^{\dagger}\mathrm{e}^{-\Psi}\Phi_{-}). \tag{12}$$

Parity conservation requires that Ψ transform as a pseudoscalar super-field.

To construct a gauge-invariant Lagrangian for Ψ it is useful to define first the vector super-field

$$V_{\mu} = \tfrac{1}{2}(C^{-1}\gamma_{\mu}(1+\mathrm{i}\gamma_{5})/2)^{\alpha\beta}D_{\alpha}(\mathrm{e}^{-\Psi}D_{\beta}\mathrm{e}^{\Psi}). \tag{13}$$

Corresponding to (11), one finds

* For example, $(\bar{D}D)^{2}D_{\alpha} = -2\mathrm{i}\bar{D}D(\gamma_{\mu}D)_{\alpha}\partial/\partial x_{\mu}$.

‡ A detailed discussion of the properties of super-fields and their use in constructing Lagrangians is in preparation.

Volume 51B, number 4 PHYSICS LETTERS 19 August 1974

$$V_\mu \to e^{i\Lambda_+} V_\mu e^{-i\Lambda_+} + i^{-1} e^{i\Lambda_+} \partial_\mu e^{-i\Lambda_+} , \quad (14)$$

i.e. a Yang-Mills type of transformation law. However, the inhomogeneous term here is an irreducible left-type field and so is annhilated by the right-handed covariant derivative (5), i.e.

$$D_{R\alpha} V_\mu \to e^{i\Lambda_+} D_{R\alpha} V_\mu e^{-i\Lambda_+} . \quad (15)$$

Also note that

$$D_{R\alpha} D_{R\beta} V_\mu \equiv 0 . \quad (16)$$

Using (15) and (16) it is easy to see that a suitable Lagrangian for the gauge field is given by

$$\mathrm{Tr}\, \bar{D} D (C^{-1})^{\alpha\beta} ((D_{R\alpha} V_\mu)(D_{R\beta} V_\mu) + (D_{L\alpha} V_\mu^\dagger)(D_{L\beta} V_\mu^\dagger))$$

$$= \tfrac{1}{2}\, \mathrm{Tr}\, (\bar{D} D)^2 (V_\mu V_\mu + V_\mu^\dagger V_\mu^\dagger) , \quad (17)$$

when a surface term is disgarded.

In a general gauge the Lagrangian (17) with V_μ defined by (13) is not a polynomial in the field variables and therefore the S matrix is not manifestly renormalizable. However, there does exist a remarkable gauge (due to Wess and Zumino [4]) in which the Lagrangian assumes a manifestly renormalizable form. To lowest order in Λ and Ψ the transformation rule (11) reads

$$\Psi \to \Psi - i\Lambda_+ + i\Lambda_- + \ldots \quad (18)$$

This indicates that from among the sixteen components in Ψ we can gauge away eight and leave Ψ in the special form:

$$\Psi = \tfrac{1}{4}\bar{\theta} i\gamma_\nu \gamma_5 \theta A_\nu + \tfrac{1}{4}\bar{\theta}\theta\bar{\theta}\gamma_5 \lambda + \tfrac{1}{32}(\bar{\theta}\theta)^2 D , \quad (19)$$

where A_ν is a transverse vector ($\partial_\nu A_\nu = 0$). In this gauge, V_μ reduces to a polynomial in Ψ,

$$V_\mu = \bar{D}\gamma_\mu \frac{1+i\gamma_5}{2} D\Psi$$

$$- \tfrac{1}{2}\bar{D}\Psi\gamma_\mu D\Psi - \tfrac{1}{2}\left[\Psi, \bar{D}\gamma_\mu \frac{1+i\gamma_5}{2} D\Psi\right] \quad (20)$$

and the Lagrangian (17) reduces to

$$\mathrm{Tr}\, [-\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu])^2$$

$$+ \bar{\lambda} i\gamma_\mu (\partial_\mu \lambda + i[A_\mu, \lambda]) + \tfrac{1}{2} D^2] . \quad (21)$$

Apart from the $\frac{1}{2}D^2$ term (which on variation gives the field equation $D = 0$ in the absence of matter), this is the Yang-Mills Lagrangian for the gauge fields A_μ and the Majorana spinors λ. The super-symmetry implies that each of the fields A_μ and λ acts as a gauge particle for the other. It would be interesting to investigate if this system (with a Majorana λ) shows the same diminution of infinities as has been noticed by Illiopoulos and Zumino [6]) for other super-symmetric Lagrangians. For example, it is amusing that charge renormalization in this theory is finite in the lowest order provided there are in the theory three distinct matter super-multiplets (each containing two spin-zero bosons and a Majorana fermion) belonging to the adjoint representation of the internal symmetry group. Whether on account of super-invariance of the theory such a result persists in higher orders is an open question.

References

[1] J. Wess and B. Zumino, Nucl. Phys. B70 (1974) 39.

[2] J-L. Gervais and B. Sakita, Nucl. Phys. B34 (1971) 632; A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86; P. Ramond, Phys. Rev. D3 (1971) 2415; Y. Aharonov, A. Casher and L. Susskind, Phys. Letters 35B (1971) 512; Y. Iwasaki and K. Kikkawa, City College (NY) preprint (1973).

[3] J. Wess and B. Zumino, Phys. Letters 49B (1974) 52.

[4] J. Wess and B. Zumino, CERN preprint TH.1857 (1974), submitted to Nucl. Phys. B.

[5] A. Salam and J. Strathdee, ICTP, Trieste, preprint IC/74/11, to be published in Nucl. Phys. B; C. Fronsdal, ICTP, Trieste, preprint IC/74/21; S. Ferrara, J. Wess and B. Zumino, CERN preprint TH.1863 (1974).

[6] J. Illiopoulos and B. Zumino, CERN preprint TH.1834 (1974), to be published in Nucl. Phys. B.

PHYSICAL REVIEW D VOLUME 11, NUMBER 6 15 MARCH 1975

Superfields and Fermi-Bose symmetry

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England

J. Strathdee
International Centre for Theoretical Physics, Trieste, Italy

(Received 12 June 1974)

The realization of supergauge transformations on fields defined over an 8-dimensional space whose points are labeled by x_μ and the anticommuting Majorana spinor θ_α is described. The covariant derivative is defined and applied to the problem of decomposing superfields into irreducible (chiral) parts and to the problem of constructing "supersymmetric" Lagrangians. Further, it is shown how to build internal symmetries (both global and local) into these Lagrangians. An example is discussed in which the internal (global) symmetry is spontaneously violated, giving rise to a supermultiplet of Goldstone particles (including fermions). When a local symmetry is broken the Higgs mechanism (for bosons and fermions) is shown to be operative. A possible solution to the problem of defining a conserved fermion number is indicated.

I. INTRODUCTION

The concept of a fundamental symmetry between fermions and bosons has begun recently to receive a good deal of attention. This symmetry was formulated first in the context of dual model theory.[1] In this 2-dimensional setting it takes the form of a *local* symmetry and plays a vital role in the elimination of ghosts. More recently, Wess and Zumino[2,3] took the decisive step of formulating a *global* Fermi-Bose symmetry in 4-dimensional spacetime.

An approach to the problem of implementing the global Fermi-Bose supersymmetry, which is somewhat different from that of Wess and Zumino, was proposed by ourselves.[4,5] This involved the consideration of a superfield $\Phi(x, \theta)$ defined on an 8-dimensional space which is the product of ordinary spacetime with a 4-dimensional space whose points are labeled by the anticommuting Majorana spinor θ_α.[6] The purpose of this article is to discuss in more detail the properties of such superfields and their products and to show how it is possible to set up supersymmetric Lagrangians which are compatible with local internal symmetries. The plan of the article is as follows.

In Sec. II a pseudogeometrical point of view is introduced and the action of the supersymmetry group on the space of x_μ and θ_α is defined. This leads in a natural way to the transformation rules for scalar, spinor, etc. superfields. The structure of such representations is discussed briefly. The important concept of covariant differentiation is introduced in Sec. III and its use in decomposing superfields into irreducible pieces is indicated. Detailed properties of the covariant derivative, including a number of identities, are set out in Appendix A. In Sec. IV the Lagrangian for a self-coupled scalar superfield first exhibited by Wess and Zumino[3] is given as an illustration. Section V is devoted to the problem of incorporating internal symmetries in a supersymmetric scheme. Those of the global kind are easily fitted in and so are mentioned only in passing. The main problem is to set up local (and particularly non-Abelian) internal symmetries which are compatible with supersymmetry. This is nontrivial. In particular, it is found that zero-mass "gauge" spinors must accompany the usual vector gauge fields.[7] Goldstone and Higgs mechanisms are discussed in Sec. VI and Goldstone fermions are, shown to arise when internal symmetry is spontaneously broken.

One of the disturbing features of the supersymmetry scheme is the prevalence of Majorana spinors. On the face of it, in combining fermions and bosons into a single multiplet it would seem to be inevitable that quantum numbers such as electric charge or baryon number must be shared between the fermions and the bosons in the multiplet, while neutral bosons would go together with neutral (i.e., Majorana) fermions. Fortunately, this difficulty is not inevitable. A counterexample is given at the end of Sec. V, where it is shown that if the system of gauge fields is allowed to interact with a massless scalar multiplet [belonging to the adjoint representation of an internal symmetry, e.g., SU(n)] then a new symmetry appears. The Majorana spinors from the gauge system can be combined with the spinors from the matter supermultiplets into *complex* Dirac spinors, with the Lagrangian exhibiting a fermion-number conservation.

With the incorporation of local internal symmetries into the framework of supersymmetric

Lagrangians and with fermion-number conservation implementable, the supersymmetric Lagrangian concept is now sufficiently developed to be considered for use in a realistic unified theory of weak, electromagnetic, and strong interactions.

II. SUPERFIELDS AND THEIR TRANSFORMATIONS

Superfields are defined over the 8-dimensional space whose points are represented by the pair (x_μ, θ_α), where x_μ denotes the usual (real) space-time coordinate and θ_α is a Majorana spinor.[8] The variables θ_α differ radically from coordinates of the usual sort in that they anticommute,

$$\theta_\alpha \theta'_\beta + \theta'_\beta \theta_\alpha = 0 .$$

This has the important consequence that any *local* function $f(\theta)$ must be a polynomial. This can be seen from the fact that the monomials

$$\theta_{\alpha_1} \theta_{\alpha_2} \cdots \theta_{\alpha_n}$$

must be antisymmetric and therefore vanishing for $n > 4$. The local function $f(\theta)$ is fully specified by 16 elements: the coefficients in its expansion in powers of θ. Half of these elements are numbers of the ordinary sort while the rest are anticommuting quantities. One can think of the function $f(\theta)$ as a kind of 16-vector. Nevertheless, it is very useful to view such objects as functions defined over a 4-space, and this aspect will be emphasized in the following.

The action of the Poincaré group on the space of x and θ is given by

$$\begin{aligned} x_\mu &\to \Lambda_{\mu\nu} x_\nu + b_\mu , \\ \theta_\alpha &\to a_\alpha{}^\beta(\Lambda)\theta_\beta , \end{aligned} \tag{2.1}$$

where $a(\Lambda)$ denotes the Dirac spinor representation of the homogeneous Lorentz transformation Λ. In particular, space reflections are associated with the mapping

$$\theta_\alpha \to i(\gamma_0 \theta)_\alpha . \tag{2.2}$$

The factor i is necessary here for compatibility with the Majorana constraint on θ.

The action of a supergauge transformation on the space of x and θ is defined by

$$\begin{aligned} x_\mu &\to x_\mu + \tfrac{1}{2} i \bar{\epsilon} \gamma_\mu \theta , \\ \theta_\alpha &\to \theta_\alpha + \epsilon_\alpha , \end{aligned} \tag{2.3}$$

where the parameter ϵ_α must, of course, be an anticommuting Majorana spinor. The group property of the mappings (2.1) and (2.3) is easily verified. However, it should perhaps be emphasized that our constructions are purely formal. For example, the spacetime translation $\frac{1}{2} i \bar{\epsilon} \gamma_\mu \theta$ is *not* a set of four ordinary real numbers such as x_μ is usually taken to be. These numbers are nilpotent, $(\bar{\epsilon}\gamma_\mu\theta)^5 = 0$. For consistency we should regard the coordinates x_μ also as belonging to some non-trivial algebra. Perhaps it may be possible to establish a rigorous geometry[9] on the space of x and θ.

The scalar superfield is naturally defined as one which transforms according to

$$\Phi'(x', \theta') = \Phi(x, \theta) . \tag{2.4}$$

Generalization to spinor and tensor superfields is equally natural. For example, the spinor would transform according to

$$\Psi'_\alpha(x', \theta') = a_\alpha{}^\beta(\Lambda)\Psi_\beta(x, \theta) .$$

As remarked above, any local function of θ must be a polynomial. To illustrate we give the expansion of the scalar superfield,

$$\begin{aligned} \Phi(x, \theta) = {} & A(x) + \bar{\theta}\psi(x) + \tfrac{1}{4}\bar{\theta}\theta F(x) + \tfrac{1}{4}\bar{\theta}\gamma_5\theta G(x) \\ & + \tfrac{1}{4}\bar{\theta} i\gamma_\nu\gamma_5\theta A_\nu(x) + \tfrac{1}{4}\bar{\theta}\theta\bar{\theta}\chi(x) \\ & + \tfrac{1}{32}(\bar{\theta}\theta)^2 D(x) , \end{aligned} \tag{2.5}$$

where the coefficients A, F, G, A_ν, D are ordinary Bose fields, and ψ and χ are Fermi fields. The behavior of these components under the action of the Poincaré group is clear: A, F, and D are scalars, G is a pseudoscalar, A_μ is an axial vector, ψ and χ are Dirac spinors. (The intrinsic parities are reversed in the case of a pseudoscalar superfield.) These components are all complex in general. However, it is possible to impose a reality condition on the superfield,

$$\Phi(x, \theta)^* = \Phi(x, \theta) ,$$

where the complex conjugation is understood to reverse the order of anticommuting factors. The real scalar superfield has Bose components which are real and Fermi components which are Majorana spinors.

The behavior of the component fields under the action of an infinitesimal supergauge transformation is easily deduced from (2.3) and (2.4),

$$\delta\Phi(x, \theta) = \bar{\epsilon}^\alpha \left(\frac{\partial \Phi}{\partial \bar{\theta}^\alpha} + \frac{i}{2} (\gamma_\mu \theta)_\alpha \frac{\partial \Phi}{\partial x_\mu} \right) , \tag{2.6}$$

by substituting the expansion (2.5). One finds

$$\begin{aligned} \delta A &= \bar{\epsilon}\psi , \\ \delta\psi &= \tfrac{1}{2}(F + \gamma_5 G + i\gamma_\mu\gamma_5 A_\mu - i\not{\partial} A)\epsilon , \\ \delta F &= \tfrac{1}{2}\bar{\epsilon}\chi - \tfrac{1}{2} i \bar{\epsilon}\not{\partial}\psi , \\ \delta G &= \tfrac{1}{2}\bar{\epsilon}\gamma_5\chi - \tfrac{1}{2} i \bar{\epsilon}\gamma_5\not{\partial}\psi , \\ \delta A_\nu &= \tfrac{1}{2}\bar{\epsilon} i\gamma_\nu\gamma_5\chi + \tfrac{1}{2} i \bar{\epsilon}\gamma_\mu i\gamma_\nu\gamma_5\partial_\mu\psi , \\ \delta\chi &= \tfrac{1}{2}(D - i\not{\partial}F - i\not{\partial}\gamma_5 G - i\gamma_\nu\gamma_5 i\not{\partial} A_\nu)\epsilon , \\ \delta D &= -i\bar{\epsilon}\not{\partial}\chi . \end{aligned} \tag{2.7}$$

In certain circumstances (which will be discussed in Sec. III) this representation turns out to be reducible.

A better insight into the structure of representations of the supersymmetry can be gained from an examination of the infinitesimal algebra. To obtain this algebra one represents the action of an infinitesimal transformation on $\Phi(x, \theta)$ by a commutator,

$$\frac{1}{i}\delta\Phi(x,\theta)=[\Phi(x,\theta),\bar{\epsilon}S]\,, \tag{2.8}$$

where the components of the generator S_α themselves comprise a Majorana spinor,

$$(\bar{\epsilon}S)^\dagger=\bar{\epsilon}S\,. \tag{2.9}$$

By considering the action of two infinitesimal transformations applied in succession and making use of the Jacobi identity one arrives at the consistency condition

$$[\bar{\epsilon}_1 S, \bar{\epsilon}_2 S]=\bar{\epsilon}_1\gamma_\mu\epsilon_2 P_\mu\,, \tag{2.10}$$

where P_μ is the generator of space translations. In fact it is clear from the rules (2.3) that the commutator of two supergauge transformations is a translation. At this point it is necessary to assume that the infinitesimal parameters ϵ anticommute with the generators S. From (2.10) one then extracts the anticommutation relation

$$\{S_\alpha, S_\beta\}=-(\gamma_\mu C)_{\alpha\beta}P_\mu \tag{2.11}$$

(where C is the charge-conjugation matrix). Since the matrices $\gamma_\mu C$ are symmetric[8] one sees that the left-hand side of (2.11) must be an anticommutator and hence the necessity of assuming that ϵ anticommutes with S.

The rest of the infinitesimal algebra is deduced in the same way. In addition to the usual rules for the commutators among generators of Poincaré transformations one finds

$$\begin{aligned}&[S_\alpha, P_\mu]=0\,,\\ &[S_\alpha, J_{\mu\nu}]=\tfrac{1}{2}(\sigma_{\mu\nu}S)_\alpha\,,\end{aligned} \tag{2.12}$$

indicating that S_α transforms like a Dirac spinor.

The rules (2.11) and (2.12) are fundamental. The construction of irreducible representations of this "algebra" and the development of rules for decomposing their products are the central problems of supersymmetry theory. We have chosen to regard this system as the infinitesimal algebra of a continuous group of point transformations in a space, some of whose coordinates are anticommuting c numbers. However, in the absence of a rigorous geometry in the space of x and θ this point of view is no more than a suggestive guide for one's intuition.

The construction of representations begins with the observation that supergauge transformations must leave invariant the manifold of states with fixed 4-momentum since S_α commutes with P_μ. On such a manifold the anticommutator (2.11) becomes a fixed set of numbers and we see that the operators S_α generate a *Clifford algebra*. Since this algebra has just 16 independent members, its one and only finite-dimensional irreducible representation is in terms of 4×4 matrices.[10] (There must exist infinite-dimensional representations as well, but we shall not consider them here.) The manifold of states with fixed 4-momentum is therefore reduced by the action of supergauge transformations into 4-dimensional invariant subspaces. However, these subspaces are not in general left invariant by the Wigner rotations. These transformations reduce the manifold into $(2\mathcal{J}+1)$-dimensional invariant subspaces. The subspaces which are invariant with respect to *both* supergauge and Wigner transformations are $4(2\mathcal{J}+1)$-dimensional.

The construction of unitary irreducible representations of the algebra (2.11) and (2.12) by Wigner's method has been treated elsewhere.[5] These representations are characterized by a mass, a spin ($\mathcal{J}$), and an intrinsic parity (η). Included in one of these representations are *four* irreducible representations of the Poincaré group. The $(\text{spin})^{\text{parity}}$ content is $(\mathcal{J}-\frac{1}{2})^\eta$, $\mathcal{J}^{i\eta}$, $\mathcal{J}^{-i\eta}$, $(\mathcal{J}+\frac{1}{2})^\eta$, where η takes one of the values $\pm i$ (for integer $\mathcal{J}$) or ± 1 (for half-integer $\mathcal{J}$). The rest mass is common. [The lightlike representations are 4-dimensional with helicity content $\pm\lambda$ and $\pm(\lambda+\frac{1}{2})$ with fixed λ.]

We sketch very briefly a method for constructing irreducible multiplets of fields (nonunitary representations). In a basis where γ_5 is diagonal the generators S_α become a pair of chiral spinors S_A and $S_{\dot{A}}$ $(A=1,2)$ which satisfy the algebra

$$\{S_A, S_B\}=0,\quad \{S_{\dot{A}}, S_{\dot{B}}\}=0,\quad \{S_A, S_{\dot{B}}\}=i\partial_{A\dot{B}}\,. \tag{2.13}$$

We shall treat S_A (or $S_{\dot{A}}$) as "raising operators" and $S_{\dot{A}}$ (or S_A) as "lowering operators." Let the lowest component $U(x)$ belong to some finite-dimensional representation $D(j_1, j_2)$ of the proper Lorentz group. Successive applications of S_A generate two new sets of components in the multiplet,

$$\begin{aligned}&S_A U(x)=M_A(x)\,,\\ &S_A S_B U(x)=\epsilon_{AB}V(x)\end{aligned} \tag{2.14}$$

(where ϵ_{AB} denotes the permutation symbol in two dimensions) *and no more* since the product of three undotted operators must vanish by (2.13). The complete table is then easily obtained:

$$S_A U(x) = M_A(x), \quad S_{\dot A} U(x) = 0,$$
$$S_A M_B(x) = \epsilon_{AB} V(x), \quad S_{\dot A} M_B(x) = -i\partial_{\dot A B} U(x), \quad (2.15)$$
$$S_A V(x) = 0, \quad S_{\dot A} V(x) = -i\partial_{\dot A}{}^B M_B(x).$$

A similar table [the parity transform of (2.15)] is obtained by treating $S_{\dot A}$ as the raising operator. Space reflections are incorporated by combining the two multiplets.[11]

Representations of the type (2.15) which have dimensionality $4(2j_1+1)(2j_2+1)$ are in general reducible. Thus, if $j_1 \geq j_2$ one can reduce the lowest component $U(x)$ to a tensor in $D(j_1-j_2, 0)$ by contraction with $2j_2$ powers of the operator $\partial/\partial x_\mu$. On this component one then constructs a representation of $4[2(j_1-j_2)+1]$ dimensions. This contraction is nothing more than the supersymmetry analog of, say, separating the longitudinal part $\partial_\mu V_\mu$ from a vector V_μ. In setting up a local action principle, one usually finds it necessary to employ fields which are reducible in this sense.

To conclude this section, it may be worth pointing out that the Majorana constraint is very potent in limiting the size of multiplets. If one were to treat S_α and $\bar S^{\dot\alpha}$ as independent generators, for example, and replace (2.11) by the set

$$\{S_\alpha, S_\beta\} = 0, \quad \{\bar S^{\dot\alpha}, \bar S^{\dot\beta}\} = 0, \quad \{S_\alpha, \bar S^{\dot\beta}\} = (\gamma_\mu)_\alpha{}^{\dot\beta} P_\mu,$$

then the fundamental representation would have 16 rather than 4 dimensions. It would include vector as well as scalar and spinor components. The same thing happens when the generators are generalized so as to carry an internal quantum number such as isospin. This type of generalization was discussed in Ref. 5.

III. THE COVARIANT DERIVATIVE

Although it is not easy to define an integral over θ space, there is certainly no problem with differentiation. The ordinary derivative is defined by

$$f(\theta + \delta\theta) = f(\theta) + \delta\bar\theta^{\dot\alpha} \frac{\partial f}{\partial \bar\theta^{\dot\alpha}},$$

where the infinitesimal $\delta\bar\theta$ stands to the left of $\partial f/\partial\bar\theta$ (since these quantities may anticommute it is important to fix their order).

An important role is played in the following by the differential operator

$$D_\alpha = \frac{\partial}{\partial\bar\theta^{\dot\alpha}} - \frac{i}{2}(\gamma_\mu\theta)_\alpha \frac{\partial}{\partial x_\mu}, \quad (3.1)$$

which we shall call the covariant derivative. This operator is covariant in the sense that it transforms as a Dirac spinor under Lorentz transformations and as an invariant with respect to the supergauge transformations (2.6).

The covariant derivative has the usual properties of a differential operator with two significant exceptions. Firstly, when applied to the product of two superfields its effects are distributed according to the rule

$$D_\alpha(\Phi_1\Phi_2) = (D_\alpha\Phi_1)\Phi_2 \pm \Phi_1(D_\alpha\Phi_2),$$

where the + (−) sign applies when Φ_1 is bosonic (fermionic), i.e., when $\delta\bar\theta$ commutes (anticommutes) with Φ_1. Secondly, the covariant derivatives neither commute nor anticommute. Their anticommutator is given by

$$\{D_\alpha, D_{\dot\beta}\} = -(\gamma_\mu C)_{\alpha\dot\beta} i \frac{\partial}{\partial x_\mu}. \quad (3.2)$$

The operator D_α is basically a Majorana spinor. For this reason it is useful to define another form

$$\bar D^{\dot\alpha} = (C^{-1})^{\dot\alpha\beta} D_\beta, \quad (3.3)$$

which is nothing more than a relabeling of components. The basic anticommutator can then be given in three equivalent forms:

$$\{D_\alpha, D_{\dot\beta}\} = -(\not\partial C)_{\alpha\dot\beta},$$
$$\{D_\alpha, \bar D^{\dot\beta}\} = (\not\partial)_\alpha{}^{\dot\beta}, \quad (3.4)$$
$$\{\bar D^{\dot\alpha}, \bar D^{\dot\beta}\} = (C^{-1}\not\partial)^{\dot\alpha\dot\beta}.$$

(It is often more convenient to work in momentum space.) The operators D_α clearly generate a Clifford algebra which is isomorphic to the superalgebra (2.11).

In view of the algebraic structure (3.4), only 16 independent operators can be made from products of the D_α. The most useful set is

$$1, D_\alpha, \bar D D, \bar D\gamma_5 D, \bar D i\gamma_\mu\gamma_5 D, \bar D D D_\alpha, (\bar D D)^2.$$

A product containing five or more factors must inevitably reduce. For example,

$$(\bar D D)^2 D_\alpha = -2\bar D D(\not D D)_\alpha.$$

Two other important identities are

$$\bar D\gamma_\mu D = 2p_\mu$$

and (3.5)

$$\bar D\sigma_{\mu\nu} D = 0.$$

A number of useful identities and multiplication rules are given in Appendix A.

It was mentioned before that the scalar superfield (2.5) is, in a sense, reducible. A reduction can be effected by imposing the condition

$$(\tfrac{1}{2}(1+i\gamma_5)D)_\alpha \Phi \equiv D_{L\alpha}\Phi = 0. \quad (3.6)$$

These linear differential equations are manifestly covariant (excluding space reflections). The general solution of (3.6) involves eight independent (real) components in contrast to the 16 of a general scalar superfield. This solution can be expressed

in the form

$$\Phi_-(x,\theta)=\exp(\tfrac{1}{4}\bar\theta\not\partial\gamma_5\theta)$$
$$\times\left(A_-(x)+\bar\theta\psi_-(x)+\tfrac{1}{2}\bar\theta\,\frac{1-i\gamma_5}{2}\,\theta F_-(x)\right), \quad (3.7)$$

where A_- and F_- are complex boson fields and ψ_- is a right-handed Dirac spinor, $i\gamma_5\psi_-=-\psi_-$. Likewise, the differential equations

$$D_{R\alpha}\Phi\equiv(\tfrac{1}{2}(1-i\gamma_5)D)_\alpha\Phi=0 \quad (3.6')$$

serve to define a left-handed superfield,

$$\Phi_+(x,\theta)=\exp(-\tfrac{1}{4}\bar\theta\not\partial\gamma_5\theta)$$
$$\times\left(A_+(x)+\bar\theta\psi_+(x)+\tfrac{1}{2}\bar\theta\,\frac{1+i\gamma_5}{2}\,\theta F_+(x)\right). \quad (3.7')$$

It is possible (although not necessary) to identify Φ_- with the complex conjugate of Φ_+, i.e.,

$$A_-=A_+^*,\ \psi_-=\psi_+^c,\ F_-=F_+^*. \quad (3.8)$$

(Thus ψ_+ and ψ_- are identified as the left- and right-handed components, respectively, of a Majorana spinor.)

The components of the chiral superfields behave under an infinitesimal supergauge transformation according to

$$\delta A_\pm=\bar\epsilon\psi_\pm,$$
$$\delta\psi_\pm=\frac{1\pm i\gamma_5}{2}(F_\pm-i\not\partial A_\pm)\epsilon, \quad (3.9)$$
$$\delta F_\pm=-\bar\epsilon i\not\partial\psi_\pm.$$

These representations are irreducible.

A general scalar superfield contains another (nonchiral) piece Φ_1 which is singled out by means of nonlinear differential conditions

$$\bar D\,\frac{1\pm i\gamma_5}{2}\,D\Phi_1=0. \quad (3.10)$$

The resolution

$$\Phi=\Phi_++\Phi_-+\Phi_1 \quad (3.11)$$

is effected by the projection operators

$$E_+=-\frac{1}{\partial^2}\bar D\,\frac{1-i\gamma_5}{2}\,D\bar D\,\frac{1+i\gamma_5}{2}\,D,$$
$$E_-=-\frac{1}{\partial^2}\bar D\,\frac{1+i\gamma_5}{2}\,D\bar D\,\frac{1-i\gamma_5}{2}\,D, \quad (3.12)$$
$$E_1=1+\frac{1}{4\partial^2}(\bar DD)^2.$$

In terms of the components (2.5), the resolution (3.11) takes the explicit form

$$A=A_++A_-+A_1,$$
$$\psi=\psi_++\psi_-+\psi_1,$$
$$F=F_++F_-,$$
$$G=iF_+-iF_-, \quad (3.13)$$
$$A_\mu=i\partial_\mu A_+-i\partial_\mu A_-+A_{1\mu},$$
$$\chi=-i\not\partial\psi_+-i\not\partial\psi_-+i\not\partial\psi_1,$$
$$D=-\partial^2A_+-\partial^2A_-+\partial^2A_1,$$

where $A_{1\mu}$ is transverse, $\partial_\mu A_{1\mu}=0$. This decomposition is useful in constructing local field theories if the resulting components are local fields. Such is the case only when the original components χ and D in (2.5) are themselves first and second derivatives, respectively, of local fields. If this is so we may say that the superfield is "locally reducible." Otherwise, it is not.

Because of their being defined by a linear differential condition, (3.6) or (3.6′), the chiral superfields are closed under multiplication. That is, we have

$$\Phi_1{}_+\Phi_2{}_+=\Phi_3{}_+$$
and (3.14)
$$\Phi_1{}_-\Phi_2{}_-=\Phi_3{}_-.$$

This remarkable property is very important in the construction of Lagrangians. The mixed product $\Phi_1{}_+\Phi_2{}_-$, on the other hand, is a general superfield, one which is not locally reducible. Details are given in Appendix B.

The spinor superfield $\Psi_\alpha=D_\alpha\Phi_+$ is left-handed with respect to the spinor index α, but is of the nonchiral (Φ_1) type in the complexion of its component fields. Conversely, the spinor $\Psi_\alpha=D_\alpha\Phi_1$ is a mixture of left- and right-handed chiral superfields (see Appendix B).

IV. A SIMPLE LAGRANGIAN

To illustrate the application of the superfield notation in a simple dynamical system we consider here the Lagrangian given by Wess and Zumino[3] for a scalar multiplet. This Lagrangian, $\mathcal{L}(\Phi_+,\Phi_-)$, must itself transform as a scalar superfield. In order that the equations of motion should be supercovariant, the action integral must be an invariant,

$$\delta\int dx\,\mathcal{L}=\int dx\,\bar\epsilon\left(\frac{\partial}{\partial\bar\theta}+\frac{i}{2}(\gamma_\mu\theta)\frac{\partial}{\partial x_\mu}\right)\mathcal{L}$$
$$=\bar\epsilon\,\frac{\partial}{\partial\bar\theta}\int dx\,\mathcal{L}+\text{surface term}$$
$$=0.$$

The action will be invariant (up to a variationally insignificant surface term) if it is independent of

θ. This means that every θ-dependent term in $\mathcal{L}(\Phi_+, \Phi_-)$ must have the form of a spacetime divergence. *Such a form can always be achieved by applying the covariant operator D_α a sufficient number of times: twice on a chiral or locally reducible superfield, four times on a general superfield.*

Consider the Lagrangian

$$\mathcal{L} = \tfrac{1}{8}(\bar{D}D)^2(\Phi_+\Phi_-) - \tfrac{1}{2}\bar{D}D[V(\Phi_+) + V(\Phi_-)]$$
$$= \tfrac{1}{2}\bar{D}D[\tfrac{1}{4}\Phi_+\bar{D}D\Phi_- + \tfrac{1}{4}\Phi_-\bar{D}D\Phi_+ - V(\Phi_+) - V(\Phi_-)], \tag{4.1}$$

where $\Phi_- = \Phi_+^*$ and V is a smooth function (typically a polynomial[12]). The spacetime integral of this expression is certainly invariant. The variationally important part is obtained by setting $\theta = 0$. With the help of the formulas in Appendix B one finds, for this part,

$$\mathcal{L} = \partial_\mu A_+ \partial_\mu A_- + F_+F_- + \tfrac{1}{2}\bar{\psi}i\not{\partial}\psi + [V'(A_+)F_+ + V'(A_-)F_-]$$
$$- \frac{1}{2}\left(V''(A_+)\bar{\psi}\,\frac{1+i\gamma_5}{2}\,\psi + V''(A_-)\bar{\psi}\,\frac{1-i\gamma_5}{2}\,\psi\right), \tag{4.2}$$

where $V'(A) = dV/dA$, etc. The equations of motion are

$$\partial^2 A_\pm = V''(A_\mp)F_\mp - \tfrac{1}{2}V'''(A_\mp)\bar{\psi}\,\frac{1\mp i\gamma_5}{2}\,\psi,$$
$$F_\pm = -V'(A_\mp), \tag{4.3}$$
$$i\not{\partial}\psi = V''(A_+)\,\frac{1+i\gamma_5}{2}\,\psi + V''(A_-)\,\frac{1-i\gamma_5}{2}\,\psi.$$

The Lagrangian (4.2) and equations of motion (4.3) could easily be expressed in terms of real scalar fields A, F and pseudoscalars B, G defined by

$$A_\pm = \frac{1}{\sqrt{2}}(A \pm iB), \quad F_\pm = \frac{1}{\sqrt{2}}(F \pm iG),$$

but there is no great advantage to be gained by this.

According to (4.3) the vacuum expectation values must, in the tree approximation, satisfy the equations

$$0 = V''(\langle A_\pm\rangle)V'(\langle A_\mp\rangle), \tag{4.4}$$

i.e., one or other of the factors V', V'' must vanish in the vacuum. Consider the possibilities. If $\langle V'\rangle = 0$ then $\langle V''\rangle$ is identified as the (common) mass of the multiplet. On the other hand, if $\langle V''\rangle = 0$ then the fermion is massless. This is a Goldstone (fermion) solution. However, it is an unstable one since the boson field equations take the form

$$\partial^2 A_\pm = -\langle V'\rangle\langle V'''\rangle(A_\mp - \langle A_\mp\rangle)$$
$$+ \text{interaction terms},$$

indicating that one boson component must be a tachyon. This situation may be altered when quantum corrections are included.[13]

V. LOCAL SYMMETRY

It is natural to ask whether supersymmetry is compatible with internal symmetries of the usual sort, both global and local. One finds that indeed it is: Global internal symmetries can be incorporated quite easily; local symmetries are more difficult.[7]

Consider first the global case. Suppose, for example, that the superfields Φ_+ and Φ_- transform as doublets of SU(2),

$$\Phi_+(x, \theta) \to \Omega\Phi_+(x, \theta),$$
$$\Phi_-(x, \theta) \to \Omega\Phi_-(x, \theta), \tag{5.1}$$

where Ω is an SU(2) matrix (independent of x *and* θ). The Lagrangian

$$\mathcal{L}_0 = \tfrac{1}{8}(\bar{D}D)^2(\Phi_+^\dagger\Phi_+ + \Phi_-^\dagger\Phi_-)$$
$$- \tfrac{1}{2}M\bar{D}D(\Phi_-^\dagger\Phi_+ + \Phi_+^\dagger\Phi_-) \tag{5.2}$$

is both supersymmetric (up to a surface term) and SU(2)-invariant. Unfortunately, the simplest SU(2)-invariant interaction for this system,

$$\mathcal{L}_1 = g\bar{D}D[(\Phi_-^\dagger\Phi_+)^2 + (\Phi_+^\dagger\Phi_-)^2],$$

is not renormalizable. To have a renormalizable (i.e., trilinear) interaction it is necessary to bring in singlet and/or triplet superfields. Thus, if $\Phi'_\pm$ transforms according to

$$\Phi'_\pm \to \Omega\Phi'_\pm\Omega^{-1} \tag{5.3}$$

then a renormalizable and SU(2)-invariant interaction is given by

$$g\bar{D}D(\Phi_-^\dagger\Phi'_+\Phi_+ + \Phi_+^\dagger\Phi'_-\Phi_-). \tag{5.4}$$

As a second example, suppose Φ_+ is a 3×3 matrix of superfields which belongs to the (real) (3, 3) representation of SU(2)×SU(2),

$$\Phi_+ \to R_1\Phi_+R_2^T, \tag{5.5}$$

where R_1 and R_2 are orthogonal matrices. Suppose, moreover, that $\Phi_- = \Phi_+^*$ so that

$$\Phi_-^T \to R_2\Phi_-^TR_1^T. \tag{5.6}$$

Then a renormalizable and invariant Lagrangian is given by

$$\mathcal{L} = \tfrac{1}{8}(\bar{D}D)^2\,\mathrm{Tr}(\Phi_-^T\Phi_+) - \tfrac{1}{4}\bar{D}D\,\mathrm{Tr}(\Phi_+^T\Phi_+ + \Phi_-^T\Phi_-)$$
$$+ g\bar{D}D(\det\Phi_+ + \det\Phi_-). \tag{5.7}$$

This Lagrangian will be considered further in Sec. VI.

The problem of setting up a *local* symmetry which is compatible with the supersymmetry is

more interesting. First of all, if the matrix Ω which appears in (5.1) is to depend on x, then it must also depend on θ. In fact, it must be a set of chiral-type superfields,

$$\Phi_\pm(x,\theta) \to \Omega_\pm(x,\theta)\Phi_\pm(x,\theta), \tag{5.8}$$

since the transformed field must, if it is to have the same number of independent components as the original, be chiral. The consistency of (5.8) relies on the closure property (3.14) of chiral superfields under multiplication.

Representing the doublet $\Phi_+(x,\theta)$ and the matrix Ω_+ by the expansions

$$\Phi_+(x,\theta) = \exp(-\tfrac{1}{4}\bar\theta\partial\!\!\!/\gamma_5\theta) \times [A_+(x) + \bar\theta\psi_+(x) + \tfrac{1}{4}\bar\theta(1+i\gamma_5)\theta F_+(x)],$$

$$\Omega_+(x,\theta) = \exp(-\tfrac{1}{4}\bar\theta\partial\!\!\!/\gamma_5\theta) \times [U_+(x) + \bar\theta V_+(x) + \tfrac{1}{4}\bar\theta(1+i\gamma_5)\theta W_+(x)]$$

the explicit form of the transformation rule (5.8) is

$$A_+(x) \to U_+(x)A_+(x),$$

$$\psi_+(x) \to U_+(x)\psi_+(x) + V_+(x)A_+(x),$$

$$F_+(x) \to U_+(x)F_+(x) - \bar V_+(x)\psi_+(x) + W_+(x)A_+(x).$$

The matrices U_+, V_+, and W_+ are complex, and V_+ carries a Dirac spinor index.[14]

It is possible to regard Ω_- as completely independent of Ω_+ or one can impose the constraints

$$\det\Omega_\pm = 1, \quad \Omega_+^\dagger = (\Omega_-)^{-1}, \tag{5.9}$$

which we shall adopt in the following. This means that the supersymmetric mass term

$$\bar D D(\Phi_-^\dagger\Phi_+ + \Phi_+^\dagger\Phi_-) \tag{5.10}$$

is an invariant of the local symmetry.

The main problem is to construct a gauge-invariant kinetic energy for the doublets $\Phi_\pm$. The expression (5.2) is certainly not satisfactory since $\Omega_+^\dagger\Omega_+ \neq 1$. It is necessary to introduce some gauge fields. Following Wess and Zumino, who solved the problem of making supersymmetry compatible with a local U(1) symmetry,[7] we deal with the compatibility problem for the case of the local internal symmetry SU(2) [in fact, the SU(2) could be generalized to a local internal SU(n)].

The gauge field Ψ is a general (not chiral) pseudoscalar Hermitian matrix superfield which transforms under the local symmetry according to the rule

$$e^{g\Psi} \to \Omega_- e^{g\Psi}\Omega_+^{-1}. \tag{5.11}$$

With the help of this field, the kinetic term can be expressed in the invariant and supersymmetric form

$$\tfrac{1}{8}(\bar D D)^2(\Phi_+^\dagger e^{g\Psi}\Phi_+ + \Phi_-^\dagger e^{-g\Psi}\Phi_-). \tag{5.12}$$

Although this is not in general a polynomial, there does exist a special gauge in which $\Psi^n = 0$ for $n \geq 3$, where (5.12) defines a renormalizable interaction. We shall come back to this in the following.

Having introduced a gauge coupling into the system of matter fields $\Phi_\pm$, we find it necessary now to set up a gauge-invariant kinetic term for Ψ. By exploiting the chiral properties of $\Omega_\pm$,

$$((1 \mp i\gamma_5)D)_\alpha\Omega_\pm(x,\theta) = 0,$$

one can prove that the vector superfield V_μ defined by

$$V_\mu = -\frac{1}{g}\left(C^{-1}\gamma_\mu\frac{1+i\gamma_5}{2}\right)^{\alpha\beta} D_\alpha(e^{-g\Psi}D_\beta e^{g\Psi}) \tag{5.13}$$

transforms under the local symmetry like a Yang-Mills gauge field,

$$V_\mu \to \Omega_+ V_\mu \Omega_+^{-1} + \frac{2i}{g}\Omega_+\partial_\mu\Omega_+^{-1}. \tag{5.14}$$

The Hermitian conjugate transforms according to

$$V_\mu^\dagger \to \Omega_- V_\mu^\dagger \Omega_-^{-1} + \frac{2i}{g}\Omega_-\partial_\mu\Omega_-^{-1}.$$

Notice that the combinations

$$\left(\frac{1-i\gamma_5}{2}D\right)_\alpha V_\mu$$

and

$$\left(\frac{1+i\gamma_5}{2}D\right)_\alpha V_\mu^\dagger$$

transform homogeneously, i.e., like field strengths. Further, one can show that these field strengths are chiral,

$$\left(\frac{1-i\gamma_5}{2}D\right)_\alpha\left(\frac{1-i\gamma_5}{2}D\right)_\alpha V_\mu = 0 \tag{5.15}$$

(and likewise for $V_\mu^\dagger$). This means that the expression

$$-\frac{1}{32}\bar D D\,\mathrm{Tr}\left(\left(C^{-1}\frac{1-i\gamma_5}{2}\right)^{\alpha\beta}(D_\alpha V_\mu)(D_\beta V_\mu) + \left(C^{-1}\frac{1+i\gamma_5}{2}\right)^{\alpha\beta}(D_\alpha V_\mu^\dagger)(D_\beta V_\mu^\dagger)\right)$$

is an invariant of the local symmetry and is supersymmetric. It can serve as a Lagrangian for the gauge field. Making use of the identity (5.15) and discarding a variationally insignificant surface term, one can put this Lagrangian into the compact form

$$\tfrac{1}{128}(\bar D D)^2\,\mathrm{Tr}(V_\mu V_\mu + V_\mu^\dagger V_\mu^\dagger). \tag{5.16}$$

Gathering together the terms (5.10), (5.12), and (5.16) we obtain the gauge-invariant and super-

symmetric Lagrangian

$$\mathcal{L} = \tfrac{1}{8}(\overline{D}D)^2[\tfrac{1}{16}\,\mathrm{Tr}(V_\mu V_\mu + V_\mu^\dagger V_\mu^\dagger) + (\Phi_+^\dagger e^{gV}\Phi_+ + \Phi_-^\dagger e^{-gV}\Phi_-)] - \tfrac{1}{2}M\overline{D}D(\Phi_-^\dagger\Phi_+ + \Phi_+^\dagger\Phi_-), \tag{5.17}$$

where V_μ is defined by (5.13).

To write out the Lagrangian (5.17) explicitly in terms of component fields would be a complicated and unrewarding task. Fortunately, there exists a remarkable gauge[7] in which the Lagrangian assumes polynomial form. The infinitesimal form of the transformation law (5.11) is, to lowest order in Ψ,

$$\delta\Psi = \delta\Omega_- - \Psi\,\delta\Omega_+ + \frac{1}{g}\,\delta\Omega_- - \frac{1}{g}\,\delta\Omega_+ . \tag{5.18}$$

This indicates that half of the components in Ψ can be transformed away, leaving it in the special form

$$\Psi = \tfrac{1}{4}\,\overline{\theta}i\gamma_\nu\gamma_5\theta A_\nu + \frac{1}{2\sqrt{2}}\,\overline{\theta}\theta\,\overline{\theta}\gamma_5\lambda + \tfrac{1}{16}(\overline{\theta}\theta)^2 D_5, \tag{5.19}$$

where A_ν is transverse.[15] In the SU(2) space the matrices A_ν, $\overline{\theta}\gamma_5\lambda$, and D_5 are Hermitian and traceless. In the special gauge, V_μ is given by

$$V_\mu^k = \overline{D}\gamma_\mu\,\frac{1+i\gamma_5}{2}\,D\Psi^k - \frac{ig}{2}\,\epsilon^{klm}\left(\overline{D}\Psi^l\gamma_\mu D\Psi^m + 2\Psi^l\,\overline{D}\gamma_\mu\,\frac{1+i\gamma_5}{2}\,D\Psi^m\right). \tag{5.20}$$

After some tedious labor one finds

$$\tfrac{1}{128}(\overline{D}D)^2\,\mathrm{Tr}(V_\mu V_\mu + V_\mu^\dagger V_\mu^\dagger) = -\tfrac{1}{4}(\partial_\mu A_\nu^k - \partial_\nu A_\mu^k + g\epsilon^{klm}A_\mu^l A_\nu^m)^2 + \tfrac{1}{2}i\overline{\lambda}^k\gamma_\mu(\partial_\mu\lambda^k + g\epsilon^{klm}A_\mu^l\lambda^m) + \tfrac{1}{2}(D_5^k)^2, \tag{5.21}$$

i.e., the Lagrangian for a Yang-Mills field A_μ^k in interaction with a triplet of Majorana spinors λ^k. The matter terms in the Lagrangian (5.17) reduce to

$$\begin{aligned}
&\tfrac{1}{8}(\overline{D}D)^2(\Phi_+^\dagger e^{gV}\Phi_+ + \Phi_-^\dagger e^{-gV}\Phi_-) - \tfrac{1}{2}M(\overline{D}D)(\Phi_-^\dagger\Phi_+ + \Phi_+^\dagger\Phi_-)\\
&\quad = (\partial_\mu A_+^\dagger + \tfrac{1}{2}igA_+^\dagger A_\mu)(\partial_\mu A_+ - \tfrac{1}{2}igA_\mu A_+) + (\partial_\mu A_-^\dagger + \tfrac{1}{2}igA_-^\dagger A_\mu)(\partial_\mu A_- - \tfrac{1}{2}igA_\mu A_-)\\
&\qquad + F_+^\dagger F_+ + F_-^\dagger F_- + i\overline{\psi}\gamma_\mu(\partial_\mu - \tfrac{1}{2}igA_\mu)\psi + M(A_-^\dagger F_+ + A_+^\dagger F_- + F_-^\dagger A_+ + F_+^\dagger A_- - \overline{\psi}\psi) + \tfrac{1}{2}g(A_+^\dagger D_5 A_+ - A_-^\dagger D_5 A_-)\\
&\qquad + \frac{ig}{\sqrt{2}}A_+^\dagger\overline{\lambda}\,\frac{1+i\gamma_5}{2}\,\psi + \frac{ig}{\sqrt{2}}A_-^\dagger\overline{\lambda}\,\frac{1-i\gamma_5}{2}\,\psi - \frac{ig}{\sqrt{2}}\,\overline{\psi}\,\frac{1+i\gamma_5}{2}\,\lambda A_- - \frac{ig}{\sqrt{2}}\,\overline{\psi}\,\frac{1-i\gamma_5}{2}\,\lambda A_+ ,
\end{aligned} \tag{5.22}$$

in which $A_\mu = A_\mu^k\tau^k$, $\lambda = \lambda^k\tau^k$, $D_5 = D_5^k\tau^k$. One may, if this is desired, replace the chiral combinations $A_\pm$ and $F_\pm$ by definite parity combinations

$$A_\pm = \frac{1}{\sqrt{2}}(A \pm iB), \quad F_\pm = \frac{1}{\sqrt{2}}(F \pm iG)$$

(bearing in mind that A, B, F, and G are all isodoublets). For a triplet of matter fields

$$\Phi_\pm = \Phi_\pm^k\tau^k \quad [\Phi_-^k = (\Phi_+^k)^*]$$

the gauge-invariant kinetic term is

$$\begin{aligned}
\tfrac{1}{16}(\overline{D}D)^2\,\mathrm{Tr}(\Phi_- e^{gV}\Phi_+ e^{-gV}) &= (\partial_\mu A_-^k + g\epsilon^{klm}A_\mu^l A_-^m)(\partial_\mu A_+^k + g\epsilon^{klm}A_\mu^l A_+^m) + F_-^k F_+^k + \tfrac{1}{2}\overline{\psi}^k i\gamma_\mu(\partial_\mu\psi^k + g\epsilon^{klm}A_\mu^l\psi^m)\\
&\quad - g\epsilon^{klm}\left(\sqrt{2}\,A_+^k\overline{\lambda}^l\,\frac{1-i\gamma_5}{2}\,\psi^m + \sqrt{2}\,A_-^k\overline{\lambda}^l\,\frac{1+i\gamma_5}{2}\,\psi^m - iA_+^k A_-^l D_5^m\right).
\end{aligned} \tag{5.23}$$

This expression is particularly interesting because the sum of (5.23) and the gauge Lagrangian (5.21) possesses a new symmetry, viz.,

$$\begin{aligned}
\lambda^k &\to \lambda^k\cos\alpha - \psi^k\sin\alpha ,\\
\psi^k &\to \lambda^k\sin\alpha + \psi^k\cos\alpha ,
\end{aligned} \tag{5.24}$$

with all boson components treated as scalars. In other words, if we introduce the complex Dirac field

$$\chi^k = \frac{1}{\sqrt{2}}(\lambda^k + i\psi^k) \tag{5.25}$$

the fermionic part of the Lagrangian takes the form

$$\overline{\chi}^k i\gamma_\mu(\partial_\mu\chi^k + g\epsilon^{klm}A_\mu^l\chi^m) + ig\epsilon^{klm}\left(\sqrt{2}\,A_+^k\overline{\chi}^l\,\frac{1-i\gamma_5}{2}\,\chi^m + \sqrt{2}\,A_-^k\overline{\chi}^l\,\frac{1+i\gamma_5}{2}\,\chi^m\right), \tag{5.26}$$

which is manifestly invariant with respect to the phase transformations

$$\chi^k \to e^{-i\alpha}\chi^k, \quad \bar{\chi}^k \to e^{i\alpha}\bar{\chi}^k . \tag{5.27}$$

The new symmetry could be associated with a conserved quantity (such as baryon or lepton number) which is carried only by the fermions.[16] To maintain this symmetry the matter multiplet must have no mass.

Before we close this section, remark that in the one-loop approximation, contributions to the Callan-Symanzik function $\beta(g)$ for the gauge supermultiplet equals $-(g^3/16\pi^2)(3C_2(G))$, while the contribution from the matter-supermultiplet equals $+(g^3/16\pi^2)C_2(G)$, where $C_2(G)$ is the value of the quadratic Casimir operator for the adjoint representation of the internal-symmetry group. If we introduce *three* matter supermultiplets (as we do in the next section), $\beta(g)=0$ in the one-loop approximation, so that the charge renormalization of g is finite.

VI. GOLDSTONE AND HIGGS PHENOMENA

It is of prime importance to demonstrate the feasibility of spontaneous-symmetry-breaking mechanisms in supersymmetric systems. It has already been pointed out that the spontaneous breakdown of supersymmetry cannot occur (at least in the tree approximation) in the case of a self-interacting scalar multiplet. Our aim now is to show that *internal* symmetries, on the other hand, can be spontaneously violated even when they are embedded in a supersymmetric scheme. This means that it will be possible to set up renormalizable supersymmetric Lagrangian models in which the gauge particles (vector and spinor) are massive.

A suitable system on which to test for spontaneous breaking is the SU(2)×SU(2)-invariant Lagrangian (5.7). To begin with we shall treat this as a global symmetry and obtain a stable Goldstone solution which carries a residual SU(2) symmetry. (This is only one of a number of possible solutions.) Then we shall go on to consider the symmetry $SU(2)_{\text{local}}\times SU(2)_{\text{global}}$ and show that the Higgs mechanism is operative.

The scalar multiplet $\Phi_{\pm}^{ia}$ belongs to the real representation (3, 3) of SU(2)×SU(2) and it satisfies the reality condition $\Phi_{-}^{ia}=(\Phi_{+}^{ia})^*$. In terms of component fields the Lagrangian (5.7) takes the form

$$\begin{aligned}\mathcal{L} &= \partial_\mu A_-^a\cdot\partial_\mu A_+^a + F_-^a\cdot F_+^a + \tfrac{1}{2}\bar{\psi}^a\cdot i\not{\partial}\psi^a + M(A_+^a\cdot F_+^a + A_-^a\cdot F_-^a - \tfrac{1}{2}\bar{\psi}^a\cdot\psi^a)\\ &\quad + g_1\epsilon^{abc}\left(A_+^a\cdot A_+^b\times F_+^c - A_+^a\cdot\bar{\psi}^b\times\frac{1+i\gamma_5}{2}\psi^c + A_-^a\cdot A_-^b\times F_-^c - A_-^a\cdot\bar{\psi}^b\times\frac{1-i\gamma_5}{2}\psi^c\right),\end{aligned} \tag{6.1}$$

where all the fields are 9-folds ($F_-^a\cdot F_+^a = F_-^{ia}F_+^{ia}$, etc.) and the fermion components ψ^{ia} are Majorana spinors. The equations of motion are

$$\begin{aligned}&-\partial^2 A_\pm^a + MF_\mp^a + g_1\epsilon^{abc}\left(2A_\mp^b\times F_\mp^c - \bar{\psi}^b\times\frac{1\mp i\gamma_5}{2}\psi^c\right) = 0,\\ &F_\mp^a + MA_\pm^a + g_1\epsilon^{abc}A_\pm^b\times A_\pm^c = 0,\\ &(i\not{\partial}-M)\psi^a - 2g_1\epsilon^{abc}\left(A_+^b\times\frac{1+i\gamma_5}{2}\psi^c + A_-^b\times\frac{1-i\gamma_5}{2}\psi^c\right) = 0 .\end{aligned} \tag{6.2}$$

From these equations it follows that the vacuum expectation values, in the tree approximation, must satisfy the algebraic equations

$$\begin{aligned}&M\langle A_\pm^a\rangle + g_1\epsilon^{abc}\langle A_\pm^b\rangle\times\langle A_\pm^c\rangle = -\langle F_\mp^a\rangle ,\\ &M\langle F_\mp^a\rangle + 2g_1\epsilon^{abc}\langle A_\mp^b\rangle\times\langle F_\mp^c\rangle = 0 .\end{aligned} \tag{6.3}$$

(We are of course requiring that the vacuum be Poincaré-invariant so that $\langle\partial^2 A_\pm\rangle = 0$ and $\langle\psi\rangle = 0$.) Equations (6.3) are very much simplified if we choose the matrix $\langle A_+^{ia}\rangle$ to be diagonal. [No loss of generality is implied since any one of the matrices involved here can be diagonalized by means of an SU(2)×SU(2) transformation.] The equations themselves then imply that the other matrices $\langle A_-^{ia}\rangle$, $\langle F_\pm^{ia}\rangle$ must be diagonal as well. Representing $\langle A_+\rangle$ by

$$\langle A_+^{ia}\rangle = \text{diag}(\lambda_1, \lambda_2, \lambda_3) \tag{6.4}$$

one finds

$$\begin{aligned}&\langle A_-^{ia}\rangle = \text{diag}(\lambda_1^*, \lambda_2^*, \lambda_3^*),\\ &\langle F_-^{ia}\rangle = -\text{diag}(M\lambda_1 + 2g_1\lambda_2\lambda_3, M\lambda_2 + 2g_1\lambda_3\lambda_1, M\lambda_3 + 2g_1\lambda_1\lambda_2),\end{aligned} \tag{6.5}$$

and Eqs. (6.3) reduce to the form

$$M^2\lambda_1 + 2Mg_1(\lambda_2\lambda_3 + \lambda_2\lambda_3^* + \lambda_2^*\lambda_3) + 4g_1^2\lambda_1(\lambda_2\lambda_2^* + \lambda_3\lambda_3^*) = 0, \tag{6.6}$$

and cyclic permutations thereof.

We shall not pursue the general solution of (6.6) but instead make the restrictive assumption that $\lambda_1=\lambda_2=\lambda_3$. In effect, we are selecting solutions in which a global SU(2) symmetry is preserved. With this restriction Eqs. (6.6) reduce to

$$\lambda(M+2g_1\lambda)(M+4g_1\lambda^*) = 0, \tag{6.7}$$

and there are three distinct (parity-conserving)

solutions.

The solution $\lambda = 0$ corresponds to the case where no symmetry is broken. The solution $\lambda = -M/4g_1$ gives

$$\langle A_{\pm}^{ia}\rangle = -\frac{M}{4g_1}\,\delta^{ia},\quad \langle F_{\pm}^{ia}\rangle = -\frac{M^2}{8g_1}\,\delta^{ia},$$

and corresponds to the case where the supersymmetry is broken as well as $SU(2)\times SU(2)$. This solution can be shown to be unstable: Some of the bosons turn out to be tachyons.

The solution $\lambda = -M/2g_1$ gives $\langle F_{\pm}\rangle = 0$ and so preserves the supersymmetry.[13] Indeed, by testing the propagation character of weak perturbations about this solution one finds that it is stable: The superfield Φ^{ia} breaks into three pieces [belonging to the representations $I = 0, 1, 2$ of the unbroken $SU(2)$] with the respective mass values

$$M_0 = M,\ M_1 = 0,\ M_2 = 2M. \tag{6.8}$$

The isovector piece is a Goldstone superfield. *It is perhaps worth emphasizing that the Goldstone multiplet includes a massless Majorana fermion along with the scalar and pseudoscalar bosons.* Supersymmetry is not broken, though the internal $SU(2)\times SU(2)$ is. It is the breaking of this internal symmetry which is responsible for the occurrence of the Goldstone fermions (together with Goldstone bosons).

Now consider what happens when the Lagrangian symmetry is generalized to $SU(2)_{\text{local}} \times SU(2)_{\text{global}}$. In the special gauge discussed in Sec. V the Lagrangian assumes the renormalizable form

$$\begin{aligned}\mathcal{L} = {}& -\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu + gA_\mu \times A_\nu)^2 + \tfrac{1}{2}\bar\lambda\cdot i\gamma_\mu(\partial_\mu\lambda + gA_\mu\times\lambda) + \tfrac{1}{2}D_5^2 \\ & + (\partial_\mu A_-^a + gA_\mu\times A_-^a)\cdot(\partial_\mu A_+^a + gA_\mu\times A_+^a) + \tfrac{1}{2}\bar\psi^a\cdot i\gamma_\mu(\partial_\mu\psi^a + gA_\mu\times\psi^a) \\ & - \sqrt{2}\,g\left(A_+^a\times\bar\lambda\cdot\frac{1-i\gamma_5}{2}\,\psi^a + A_-^a\times\bar\lambda\cdot\frac{1+i\gamma_5}{2}\,\psi^a\right) + igA_+^a\times A_-^a\cdot D_5 + F_-^a\cdot F_+^a + M(A_+^a\cdot F_+^a + A_-^a\cdot F_-^a - \tfrac{1}{2}\bar\psi^a\cdot\psi^a) \\ & + g_1\epsilon^{abc}\left(A_+^a\cdot A_+^b\times F_+^c - A_+^a\cdot\bar\psi^b\times\frac{1+i\gamma_5}{2}\,\psi^c + A_-^a\cdot A_-^b\times F_-^c - A_-^a\cdot\bar\psi^b\times\frac{1-i\gamma_5}{2}\,\psi^c\right),\end{aligned} \tag{6.9}$$

where the gauge condition $\partial_\mu A_\mu = 0$ is understood. Our purpose here is only to discover the excitation spectrum implicit in (6.9) when the symmetry $SU(2)\times SU(2)$ is broken down to $SU(2)$. We shall therefore take advantage of the manifest $SU(2)_{\text{local}}$ symmetry in (6.9) to change over from the renormalizable Landau gauge to the unitary gauge in which the spectrum is simplified. That is, we shall adopt the gauge condition

$$A^{[ia]} \equiv \tfrac{1}{2}(A^{ia} - A^{ai}) = 0, \tag{6.10}$$

where A^{ia} denotes the real (scalar) part of $A_{\pm}^{ia}$. The scalar isovector part of the superfield is, in effect, "gauged away." The pseudoscalar and spinor parts must of course remain.

Into (6.9) substitute

$$\sqrt{2}\,A_{\pm}^{ia} = -\frac{M}{\sqrt{2}\,g_1}\,\delta^{ia} + A^{(ia)} \pm iB^{ia} \tag{6.11}$$

and collect the bilinear terms

$$\begin{aligned}\mathcal{L}^{(2)} = {}& -\tfrac{1}{4}(\partial_\mu A_\nu^j - \partial_\nu A_\mu^j)^2 + \tfrac{1}{2}\bar\lambda^j i\not\partial\lambda^j + \tfrac{1}{2}(D_5^j)^2 + \frac{1}{2}\left(\partial_\mu A^{(ia)} - \frac{Mg}{\sqrt{2}\,g_1}\,\epsilon^{ija}A_\mu^j\right)^2 + \tfrac{1}{2}(\partial_\mu B^{ia})^2 + \tfrac{1}{2}\bar\psi^{ia} i\not\partial\psi^{ia} \\ & - \tfrac{1}{2}(F^{ia})^2 + \tfrac{1}{2}(G^{ia})^2 + M(2A^{(ia)}F^{(ia)} - A^{(ii)}F^{(jj)} - 2B^{(ia)}G^{(ia)} + B^{(ii)}G^{(jj)}) \\ & - \tfrac{1}{2}M(2\bar\psi^{(ia)}\psi\psi^{(ia)} - \bar\psi^{(ii)}\psi^{(jj)}) - \frac{Mg}{\sqrt{2}\,g_1}\,\epsilon^{ija}\bar\lambda^j\psi^{ia},\end{aligned} \tag{6.12}$$

where $\psi^{(ia)} = (\tfrac{1}{2})(\psi^{ia} + \psi^{ai})$, etc. This free Lagrangian can be separated into three independent pieces if the fields are decomposed into their $I = 0, 1, 2$ components by writing, for example,

$$\psi^{ia} = \psi_2^{(ia)} + \psi_1^{[ia]} + \tfrac{1}{3}\,\delta^{ia}\psi_0, \tag{6.13}$$

where $\psi_2^{(ii)} = 0$. Suppressing I-spin indices, the three pieces are given by

$$\begin{aligned}\mathcal{L}_0 = {}& \tfrac{1}{2}(\partial_\mu A_0)^2 + \tfrac{1}{2}(\partial_\mu B_0)^2 + \tfrac{1}{2}\bar\psi_0 i\not\partial\psi_0 + \tfrac{1}{2}(F_0^2 + G_0^2) \\ & - M(A_0F_0 - B_0G_0 - \tfrac{1}{2}\bar\psi_0\psi_0),\end{aligned} \tag{6.14}$$

$$\begin{aligned}\mathcal{L}_1 = {}& -\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 + \frac{1}{2}\left(\frac{Mg}{g_1}\right)^2 A_\mu^2 + \tfrac{1}{2}D_5^2 \\ & + \tfrac{1}{2}(\partial_\mu B_1)^2 + \tfrac{1}{2}(F_1^2 + G_1^2) \\ & + \tfrac{1}{2}\bar\psi_1 i\not\partial\psi_1 + \tfrac{1}{2}\bar\lambda i\not\partial\lambda - \frac{Mg}{g_1}\,\bar\lambda\psi_1,\end{aligned} \tag{6.15}$$

$$\begin{aligned}\mathcal{L}_2 = {}& \tfrac{1}{2}(\partial_\mu A_2)^2 + \tfrac{1}{2}(\partial_\mu B_2)^2 + \tfrac{1}{2}\bar\psi_2 i\not\partial\psi_2 + \tfrac{1}{2}(F_2^2 + G_2^2) \\ & + 2M(A_2F_2 - B_2G_2 - \tfrac{1}{2}\bar\psi_2\psi_2).\end{aligned} \tag{6.16}$$

The Lagrangians (6.14) and (6.16), respectively,

describe the propagation of an $I=0$ multiplet with mass M and an $I=2$ multiplet with mass $2M$. The Lagrangian (6.15) describes an $I=1$ system containing a massless pseudoscalar, a vector with mass Mg/g_1, and a pair of Majorana spinors,

$$\frac{1}{\sqrt{2}}(\psi_1+\lambda) \text{ and } \frac{\gamma_5}{\sqrt{2}}(\psi_1-\lambda),$$

with mass Mg/g_1 and opposite parities.

This shows that the Higgs mechanism is operating here in the usual way. The fact that a pseudoscalar Goldstone particle remains in the system merely indicates the need for a larger local symmetry: one which involves axial-vector gauge particles as well as vectors.[17] It is certainly possible to construct such a scheme by the methods described in Sec. V.

APPENDIX A: THE ALGEBRA OF COVARIANT DERIVATIVES

The differential operators D_α defined by

$$D_\alpha = \frac{\partial}{\partial\bar\theta^\alpha} - \tfrac{1}{2}(\not\partial\theta)_\alpha \tag{A1}$$

[and $\bar D^\alpha$ defined by $\bar D^\alpha = (C^{-1})^{\alpha\beta}D_\beta$] are easily seen to generate a Clifford algebra which is isomorphic to the supersymmetry algebra, viz.,

$$\{D_\alpha, D_\beta\} = -(\not\partial C)_{\alpha\beta}\ . \tag{A2}$$

The purpose of this Appendix is to list some of its properties.

First of all, the algebra contains 16 independent basis elements,

$$1,\ D_\alpha,\ \bar DD,\ \bar D\gamma_5 D,\ \bar D i\gamma_\mu\gamma_5 D,\ \bar DDD_\alpha,\ (\bar DD)^2\ . \tag{A3}$$

Any product of D's can be reduced to a linear combination of these. A complete multiplication table for these basis elements would be too bulky to reproduce here. Instead, we shall list only the more difficult products from which, with the help of (A2), any other product can be deduced without too much effort.

To begin with, the product of two D's is given by

$$D_\alpha D_\beta = -\tfrac{1}{2}(\not\partial C)_{\alpha\beta} + \tfrac{1}{4}C_{\alpha\beta}\bar DD - \tfrac{1}{4}(\gamma_5 C)_{\alpha\beta}\bar D\gamma_5 D - \tfrac{1}{4}(i\gamma_\nu\gamma_5 C)_{\alpha\beta}\bar D i\gamma_\nu\gamma_5 D\ . \tag{A4}$$

Multiplication of this formula by $-(C^{-1}\gamma_\mu)^{\alpha\beta}$ and by $-(C^{-1}\sigma_{\mu\nu})^{\alpha\beta}$ yields the identities

$$\bar D\gamma_\mu D = 2p_\mu$$

and (A5)

$$\bar D\sigma_{\mu\nu}D = 0\ .$$

Products of three D's are comprised in the formulas

$$\begin{aligned}
&D_\alpha\bar DD = \bar DDD_\alpha + (2\not\partial D)_\alpha\ ,\\
&D_\alpha\bar D\gamma_5 D = -\bar DD(\gamma_5 D)_\alpha + (2\not\partial\gamma_5 D)_\alpha\ ,\\
&D_\alpha\bar D i\gamma_\mu\gamma_5 D = -\bar DD(i\gamma_\mu\gamma_5 D)_\alpha + 2ip_\mu(\gamma_5 D)_\alpha\ ,\\
&\bar D\gamma_5 DD_\alpha = -\bar DD(\gamma_5 D)_\alpha\ ,\\
&\bar D i\gamma_\mu\gamma_5 DD_\alpha = -\bar DD(i\gamma_\mu\gamma_5 D)_\alpha - (2p_\lambda\sigma_{\lambda\mu}\gamma_5 D)_\alpha\ .
\end{aligned} \tag{A6}$$

Products of four D's can sometimes be reduced by multiplying a single D into one of the formulas (A6). For example,

$$\begin{aligned}
(\bar D\gamma_5 D)^2 &= \bar D\gamma_5 DD_\alpha(-C^{-1}\gamma_5 D)^\alpha\\
&= -\bar DD(\gamma_5 D)_\alpha(-C^{-1}\gamma_5 D)^\alpha\\
&= -\bar DD(-\bar D\gamma_5 C)_\alpha(-C^{-1}\gamma_5 D)^\alpha\\
&= (\bar DD)^2\ .
\end{aligned}$$

Any combination of four D's can be reduced with the help of (A4) and the multiplication table

	$\bar DD$	$\bar D\gamma_5 D$	$\bar D i\gamma_\nu\gamma_5 D$
$\bar DD$	$(\bar DD)^2$	$-2ip_\rho\bar D i\gamma_\rho\gamma_5 D$	$2ip_\nu\bar D\gamma_5 D$
$\bar D\gamma_5 D$	$2ip_\rho\bar D i\gamma_\rho\gamma_5 D$	$(\bar DD)^2$	$-2ip_\nu\bar DD$
$\bar D i\gamma_\mu\gamma_5 D$	$-2ip_\mu\bar D\gamma_5 D$	$2ip_\mu\bar DD$	$\eta_{\mu\nu}(\bar DD)^2 - 2ip_\lambda\epsilon_{\lambda\mu\nu\rho}\bar D i\gamma_\rho\gamma_5 D - 4(\eta_{\mu\nu}p^2 - p_\mu p_\nu)$

(A7)

where left (right) factors are listed in the rows (columns).

Products of five or more D's can be reduced with the help of the above results together with the fundamental formula

$$(\bar DD)^2 D_\alpha = -\bar DD(2\not\partial D)_\alpha\ . \tag{A8}$$

Of some importance are the identities

$$((1\pm i\gamma_5)D)_\alpha\bar D(1\pm i\gamma_5)D = 0\ , \tag{A9}$$

which become more transparent in the notation of 2-component spinors. In a basis where γ_5 is diagonal, the anticommutators (A2) take the form

$$\{D_A, D_B\} = 0,\ \{D_{\dot A}, D_{\dot B}\} = 0,\ \{D_A, D_{\dot B}\} = p_{A\dot B}\ .$$

Since they anticommute, the product of three operators D_A must vanish, e.g.,

$$0 = D_A D_B D_C = \epsilon_{BC}D_A\bar D\,\frac{1+i\gamma_5}{2}D, \text{ etc.}$$

This is one of the identities (A9).

Frequently useful are the following formulas which give explicitly the action of some operators on a general superfield Φ whose components are given in (2.5):

$$-\bar{D}\,\frac{1\pm i\gamma_5}{2}\,D\Phi = F\mp iG+\bar{\theta}\,\frac{1\mp i\gamma_5}{2}\,(\chi-i\not{\partial}\psi)$$
$$+\frac{1}{4}\,\bar{\theta}\,\frac{1\mp i\gamma_5}{2}\,\theta(D-\partial^2 A\pm 2i\partial_\mu A_\mu)$$
$$+\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5\theta(\mp i\partial_\nu)(F\mp iG)$$
$$+\tfrac{1}{4}\bar{\theta}\theta\bar{\theta}\,\frac{1\pm i\gamma_5}{2}\,(-i\not{\partial})(\chi-i\not{\partial}\psi)$$
$$+\tfrac{1}{32}\,(\bar{\theta}\theta)^2(-\partial^2)(F\mp iG)\,, \qquad \text{(A10)}$$

$$\bar{D}i\gamma_\mu\gamma_5 D\Phi = 2A_\mu+\bar{\theta}(i\gamma_\mu\gamma_5\chi-\not{\partial}\gamma_\mu\gamma_5\psi)$$
$$+\tfrac{1}{4}\bar{\theta}\theta(2\partial_\mu G)+\tfrac{1}{4}\bar{\theta}\gamma_5\theta(-2\partial_\mu F)$$
$$+\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5\theta[\eta_{\mu\nu}D-2\epsilon_{\mu\nu\lambda\rho}\partial_\lambda A_\rho$$
$$+(\eta_{\mu\nu}\partial^2-2\partial_\mu\partial_\nu)A]$$
$$+\tfrac{1}{4}\bar{\theta}\theta\bar{\theta}[\gamma_\mu\not{\partial}\gamma_5\chi-(\eta_{\mu\nu}\partial^2-2\partial_\mu\partial_\nu)i\gamma_\nu\gamma_5\psi]$$
$$+\tfrac{1}{32}\,(\bar{\theta}\theta)^2(\eta_{\mu\nu}\partial^2-2\partial_\mu\partial_\nu)A_\nu\,, \qquad \text{(A11)}$$

$$\tfrac{1}{2}(\bar{D}D)^2\Phi = D-\partial^2 A-\bar{\theta}(\partial^2\psi+i\not{\partial}\chi)-\tfrac{1}{4}\bar{\theta}\theta(2\partial^2 F)$$
$$-\tfrac{1}{4}\bar{\theta}\gamma_5\theta(2\partial^2 G)-\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5\theta(2\partial_\nu\partial_\mu A_\mu)$$
$$+\tfrac{1}{4}\,\bar{\theta}\theta\bar{\theta}(i\not{\partial})(\partial^2\psi+i\not{\partial}\chi)$$
$$-\tfrac{1}{32}(\bar{\theta}\theta)^2\partial^2(D-\partial^2 A)\,, \qquad \text{(A12)}$$

$$D_\alpha\Phi = \psi_\alpha+\tfrac{1}{2}[-i(\gamma_\mu\theta)_\alpha\partial_\mu A+\theta_\alpha F+(\gamma_5\theta)_\alpha G+(i\gamma_\nu\gamma_5\theta)_\alpha A_\nu]$$
$$+\tfrac{1}{4}\bar{\theta}\theta[\tfrac{1}{2}i(\not{\partial}\psi)_\alpha+\tfrac{1}{2}\chi_\alpha]$$
$$+\tfrac{1}{4}\bar{\theta}\gamma_5\theta[-\tfrac{1}{2}i(\not{\partial}\gamma_5\psi)_\alpha+\tfrac{1}{2}(\gamma_5\chi)_\alpha]$$
$$+\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5[-\tfrac{1}{2}i\not{\partial}(i\gamma_\nu\gamma_5\psi)_\alpha+\tfrac{1}{2}(i\gamma_\nu\gamma_5\chi)_\alpha]$$
$$+\tfrac{1}{4}\bar{\theta}\theta\tfrac{1}{2}[-i(\gamma_\mu\theta)_\alpha\partial_\mu F+(i\gamma_\mu\gamma_5\theta)_\alpha\partial_\mu G$$
$$-(\gamma_\mu\gamma_\nu\gamma_5\theta)_\alpha\partial_\mu A_\nu+\theta_\alpha D]$$
$$+\tfrac{1}{32}(\bar{\theta}\theta)^2(i\not{\partial}\chi)_\alpha\,. \qquad \text{(A13)}$$

APPENDIX B: PRODUCTS

The product of two expansions of the form (2.5) can be rearranged into

$$\Phi(1,\theta)\Phi(2,\theta) = A(1)A(2)+\bar{\theta}[A(1)\psi(2)+\psi(1)A(2)]+\tfrac{1}{4}\bar{\theta}\theta[A(1)F(2)-\bar{\psi}^c(1)\psi(2)+F(1)A(2)]$$
$$+\tfrac{1}{4}\bar{\theta}\gamma_5\theta[A(1)G(2)+\bar{\psi}^c(1)\gamma_5\psi(2)+G(1)A(2)]$$
$$+\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5\theta[A(1)A_\nu(2)+\bar{\psi}^c(1)i\gamma_\nu\gamma_5\psi(2)+A_\nu(1)A(2)]$$
$$+\tfrac{1}{4}\bar{\theta}\theta\bar{\theta}[A(1)\chi(2)+\psi(1)F(2)-\gamma_5\psi(1)G(2)+i\gamma_\nu\gamma_5\psi(1)A_\nu(2)$$
$$+\chi(1)A(2)+F(1)\psi(2)-G(1)\gamma_5\psi(2)-A_\nu(1)i\gamma_\nu\gamma_5\psi(2)]$$
$$+\tfrac{1}{32}(\bar{\theta}\theta)^2[A(1)D(2)+2F(1)F(2)+2G(1)G(2)+2A_\nu(1)A_\nu(2)$$
$$+D(1)A(2)-2\bar{\psi}^c(1)\chi(2)-2\bar{\chi}^c(1)\psi(2)]\,, \qquad \text{(B1)}$$

where ψ^c denotes the charge conjugate of ψ, i.e.,

$$(\bar{\psi}^c)^\alpha = (C^{-1})^{\alpha\beta}\psi_\beta\,.$$

The formula (B1) comprises the multiplication table for the anticommuting Majorana spinor θ and the various monomials made from it.

A particular case of (B1) is the product of left- and right-type chiral superfields,

$$\Phi_+(1,\theta)\Phi_-(2,\theta) = A_+(1)A_-(2)+\bar{\theta}[A_+(1)\psi_-(2)+\psi_+(1)A_-(2)]+\tfrac{1}{4}\bar{\theta}\theta[A_+(1)F_-(2)+F_+(1)A_-(2)]$$
$$+\tfrac{1}{4}\bar{\theta}\gamma_5\theta[-iA_+(1)F_-(2)+iF_+(1)A_-(2)]$$
$$+\tfrac{1}{4}\bar{\theta}i\gamma_\nu\gamma_5\theta[i\partial_\nu A_+(1)A_-(2)-iA_+(1)\partial_\nu A_-(2)-\bar{\psi}_+^c(1)\gamma_\nu\psi_-(2)]$$
$$+\tfrac{1}{4}\bar{\theta}\theta\bar{\theta}[-i\not{\partial}\psi_+(1)A_-(2)+i\gamma_\nu\psi_+(1)\partial_\nu A_-(2)-iA_+(1)\not{\partial}\psi_-(2)$$
$$+i\partial_\nu A_+(1)\gamma_\nu\psi_-(2)+2\psi_+(1)F_-(2)+2F_+(1)\psi_-(2)]$$
$$+\tfrac{1}{32}(\bar{\theta}\theta)^2[-\partial^2 A_+(1)A_-(2)+2\partial_\nu A_+(1)\partial_\nu A_-(2)-A_+(1)\partial^2 A_-(2)$$
$$+4F_+(1)F_-(2)+2\bar{\psi}_+^c(i\overleftarrow{\not{\partial}}-i\overrightarrow{\not{\partial}})\psi_-(2)]\,. \qquad \text{(B2)}$$

The multiplication of two left-type fields yields again a left-type field with

$$A_+(3) = A_+(1)A_+(2)\,,$$
$$\psi_+(3) = A_+(1)\psi_+(2)+\psi_+(1)A_+(2)\,, \qquad \text{(B3)}$$
$$F_+(3) = A_+(1)F_+(2)-\bar{\psi}_+^c(1)\psi_+(2)+F_+(1)A_+(2)\,.$$

Repeated application of this rule gives the components of the superfield $(\Phi_+(x,\theta))^n$, viz.,

$$\begin{aligned} A_+ &= A_+(x)^n\,, \\ \psi_+ &= nA_+(x)^{n-1}\psi_+(x)\,, \\ F_+ &= nA_+(x)^{n-1}F_+(x) - \frac{n(n-1)}{2}A_+(x)^{n-2}\overline{\psi}_+^{\,c}(x)\psi_+(x)\,. \end{aligned} \tag{B4}$$

The vector multiplet Φ_1 contained in (3.8) can be expressed in chiral form as a *spinor* superfield,

$$\begin{aligned} \Psi_{\alpha\pm} &= \left(\frac{1\pm i\gamma_5}{2}\not{\partial}D\right)_\alpha \Phi_1 \\ &= \exp(\mp\tfrac{1}{4}\overline{\theta}\not{\partial}\gamma_5\theta)\left\{\left(\frac{1\pm i\gamma_5}{2}\lambda\right)_\alpha + \overline{\theta}^\beta\left[\frac{i}{2}\left(\frac{1\pm i\gamma_5}{2}\right)_{\beta\alpha}D \mp i\left(\frac{1\pm i\gamma_5}{2}\sigma_{\mu\nu}C\right)_{\beta\alpha}A_{\mu\nu}\right] + \tfrac{1}{4}\overline{\theta}(1\pm i\gamma_5)\theta\frac{1\pm i\gamma_5}{2}i\not{\partial}\lambda\right\}\,, \end{aligned} \tag{B5}$$

where

$$\begin{aligned} D &= \partial^2 A_1\,, \\ \lambda &= \not{\partial}\psi_1\,, \\ A_{\mu\nu} &= \partial_\mu A_{1\nu} - \partial_\nu A_{1\mu}\,. \end{aligned} \tag{B6}$$

Conversely,

$$\Phi_1 = \frac{1}{2i\partial^2}\overline{D}^\alpha(\Psi_{\alpha+} + \Psi_{\alpha-})\,, \tag{B7}$$

which may be a nonlocal superfield.

APPENDIX C: SUPERFIELD EQUATIONS

In the text we have been concerned mainly with the construction of supersymmetric Lagrangians. For this task the superfield concept has provided a compact and suggestive notational framework. One may ask whether the concept can be usefully pursued and, in particular, whether it would be advantageous to set up an apparatus of Feynman rules with superfield propagators, superfield vertices, etc. Little has been done in this direction. The purpose of this Appendix is merely to sketch a preliminary idea of how such a development would proceed.

Corresponding to the Lagrangian (4.1) the superfield equations of motion are

$$\begin{aligned} \overline{D}D\Phi_+ &= 2V'(\Phi_-)\,, \\ \overline{D}D\Phi_- &= 2V'(\Phi_+)\,. \end{aligned} \tag{C1}$$

These equations can be solved perturbatively. The first step is to linearize them and work out the propagator. Make the replacement

$$V(\Phi_\pm) \to (M/2)\Phi_\pm^{\,2} + J_\pm\Phi_\pm\,, \tag{C2}$$

where $J_\pm(x,\theta)$ denote external-source distributions of the usual chiral types

$$\begin{aligned} J_\pm(x,\theta) = \exp(\mp\tfrac{1}{4}\overline{\theta}\not{\partial}\gamma_5\theta)\Big(&J_{F_\pm} - \overline{\theta}\frac{1\pm i\gamma_5}{2}J_\psi \\ &+ \tfrac{1}{4}\overline{\theta}(1\pm i\gamma_5)\theta J_{A_\pm}\Big)\,, \end{aligned} \tag{C3}$$

where the components are labeled such that $J_{A_\pm}$ acts as the source of $A_\pm$, etc. The linear inhomogeneous equations of motion

$$\begin{aligned} \overline{D}D\Phi_+ - 2M\Phi_- &= 2J_-\,, \\ \overline{D}D\Phi_- - 2M\Phi_+ &= 2J_+ \end{aligned} \tag{C4}$$

are easily solved. Making use of the identity

$$(\overline{D}D)^2\Phi_\pm = -4\partial^2\Phi_\pm\,, \tag{C5}$$

which can be deduced from the formula (A10), one finds

$$\Phi_\pm = -\frac{1}{\partial^2 + M^2}(MJ_\pm + \tfrac{1}{2}\overline{D}DJ_\pm)\,. \tag{C6}$$

This equation defines the bare propagator. In a formal sense the propagator is given by equations like

$$\langle T\Phi_+(x_1,\theta_1)\Phi_-(x_2,\theta_2)\rangle = \frac{\hbar}{i}\frac{\delta\Phi_+(x_1,\theta_1)}{\delta J_-(x_2,\theta_2)}\,,$$

but the meaning of the functional derivative here needs to be clarified. To this end one can exploit the supposed invariance of the vacuum with respect to translations and supersymmetry transformations to write

$$\begin{aligned} \langle T\Phi_+(x_1,\theta_1)\Phi_-(x_2,\theta_2)\rangle &= \langle T\Phi_+(x_1+a+\tfrac{1}{2}i\overline{\epsilon}\gamma\theta_1,\theta_1+\epsilon) \\ &\quad\times\Phi_-(x_2+a+\tfrac{1}{2}i\overline{\epsilon}\gamma\theta_2,\theta_2+\epsilon)\rangle \\ &= \langle T\Phi_+(x_1-x_2+\tfrac{1}{2}i\overline{\theta}_1\gamma\theta_2,\theta_1-\theta_2) \\ &\quad\times\Phi_-(0,0)\rangle \end{aligned} \tag{C7}$$

(on choosing $\epsilon = -\theta_2$, $a = -x_2$). Now $\Phi_-(0,0) = A_-(0)$ and one needs to evaluate only

$$\begin{aligned} \frac{i}{\hbar}\langle T\Phi_+(x,\theta)A_-(0)\rangle &= \frac{\delta\Phi_+(x,\theta)}{\delta J_{A_-}(0)} \\ &= -\frac{1}{\partial^2+M^2}\tfrac{1}{2}\overline{D}D\frac{\delta J_-(x,\theta)}{\delta J_{A_-}(0)} \\ &= -\frac{1}{\partial^2+M^2}\tfrac{1}{2}\overline{D}D \\ &\quad\times[\exp(\tfrac{1}{4}\overline{\theta}\not{\partial}\gamma_5\theta)\tfrac{1}{4}\overline{\theta}(1-i\gamma_5)\theta\delta(x)]\,, \end{aligned} \tag{C8}$$

where we have used (C3) and (C6). This formula can be simplified with the help of (A10), which implies the relation

$$-\tfrac{1}{2}\bar{D}D\Phi_- = \exp(-\tfrac{1}{4}\bar{\theta}\not{\partial}\gamma_5\theta)[F_- + \bar{\theta}(-i\not{\partial}\psi_-) + \tfrac{1}{4}\bar{\theta}(1+i\gamma_5)\theta(-\partial^2 A_-)]$$

and, therefore,

$$\frac{i}{\hbar}\langle T\Phi_+(x,\theta)A_-(0)\rangle = \frac{1}{\partial^2+M^2}\exp(-\tfrac{1}{4}\bar{\theta}\not{\partial}\gamma_5\theta)\delta(x)\,. \tag{C9}$$

The propagator (C7) now takes the form

$$\frac{i}{\hbar}\langle T\Phi_+(x_1,\theta_1)\Phi_-(x_2,\theta_2)\rangle = \exp\left([\tfrac{1}{2}\bar{\theta}_1\gamma_\mu\theta_2 + \tfrac{1}{4}(\bar{\theta}_1-\bar{\theta}_2)i\gamma_\mu\gamma_5(\theta_1-\theta_2)]i\frac{\partial}{\partial x_\mu}\right)\frac{1}{\partial^2+M^2}\,\delta(x_1-x_2)\,, \tag{C10}$$

which is a polynomial in θ_1 and θ_2. (The component propagators can be obtained by comparing coefficients.) In similar fashion one obtains

$$\frac{i}{\hbar}\langle T\Phi_+(x_1,\theta_1)\Phi_+(x_2,\theta_2)\rangle = \exp\left([\tfrac{1}{4}\bar{\theta}_1 i\gamma_\mu\gamma_5\theta_1 - \tfrac{1}{4}\bar{\theta}_2 i\gamma_\mu\gamma_5\theta_2]i\frac{\partial}{\partial x_\mu}\right)(-\tfrac{1}{4}M)\bar{\theta}_{12}(1+i\gamma_5)\theta_{12}\frac{1}{\partial^2+M^2}\,\delta(x_1-x_2)\,. \tag{C11}$$

With the bare propagators in hand one can contemplate the problem of computing scattering amplitudes. For example, one could put the equations (C1) into integral form

$$\Phi_\pm = -\frac{1}{\partial^2+M^2}[MV'_{\rm int}(\Phi_\pm) + \tfrac{1}{2}\bar{D}DV'_{\rm int}(\Phi_\mp)] \tag{C12}$$

and go on to obtain perturbative developments of the superfields. This kind of approach may, perhaps, lead to useful insights.

A different sort of Lagrangian which can be used to characterize at least the free-field behavior is given by

$$\mathcal{L} = (\bar{D}D)^2[\tfrac{1}{2}\Phi(\bar{D}D-2M)\Phi - 2J\Phi]\,, \tag{C13}$$

where Φ is a general superfield. The equations of motion are

$$(\bar{D}D-2M)\Phi = 2J\,. \tag{C14}$$

In fact, these equations are equivalent to (C4). Substitute the resolution (3.11) for Φ and J. Then, since $\bar{D}D\Phi_1 = 0$, Eq. (C14) takes the form

$$\begin{aligned}&\bar{D}D\Phi_- - 2M\Phi_+ = 2J_+\,,\\ &\bar{D}D\Phi_+ - 2M\Phi_- = 2J_-\,,\\ &-2M\Phi_1 = 2J_1\,,\end{aligned} \tag{C15}$$

and one sees that the nonchiral part, Φ_1, does not propagate. The equation (C14) can be solved directly with the help of (A6) to give

$$\Phi = -\frac{1}{\partial^2+M^2}\left[\tfrac{1}{2}\bar{D}D + \frac{1}{4M}(\bar{D}D)^2\right]J - \frac{1}{M}J\,, \tag{C16}$$

and from this the form of the 2-point function can be obtained.

The methods can be applied to other kinds of superfield. For example, the spinor superfields, $\Psi_{\alpha\pm}$, may satisfy the equations

$$(i\not{\partial}-2M)\Psi_\pm + \tfrac{1}{2}\bar{D}D\Psi_\mp = 2J_\pm\,, \tag{C17}$$

which are solved by

$$\Psi_\pm = \frac{1}{i\not{\partial}-M}\left(J_\pm + \frac{1}{2M}\bar{D}DJ_\mp\right) - \frac{1}{2M}J_\pm\,. \tag{C18}$$

These equations describe the propagation of a supermultiplet of particles of mass M and (spin)$^{\rm parity}$ content 0^+, $(\tfrac{1}{2})^1$, $(\tfrac{1}{2})^{-1}$, and 1^+. It remains to be seen whether or not there exist any renormalizable interactions for this superfield.

[1] P. Ramond, Phys. Rev. D 3, 2415 (1971); A. Neveu and J. H. Schwarz, Nucl. Phys. B31, 86 (1971); J.-L. Gervais and B. Sakita, *ibid.* B34, 632 (1971); Y. Iwasaki and K. Kikkawa, Phys. Rev. D 8, 440 (1973).

[2] J. Wess and B. Zumino, Nucl. Phys. B70, 39 (1974).

[3] J. Wess and B. Zumino, Phys. Lett. 49B, 52 (1974).

[4] Abdus Salam and J. Strathdee, Nucl. Phys. B76, 477 (1974).

[5] Abdus Salam and J. Strathdee, Nucl. Phys. B80, 499 (1974).

[6] Superfields defined over a space of 2-component spinors have been considered by C. Fronsdal, ICTP Report No. IC/74/21 (unpublished), and by S. Ferrara, J. Wess, and B. Zumino, Phys. Lett. 51B, 239 (1974). These latter authors independently defined the covariant derivative and the chiral superfields.

[7] The problem was solved for an Abelian local symmetry by J. Wess and B. Zumino, Nucl. Phys. B78, 1 (1974), who invented the special gauge used in the text. The non-Abelian case was solved by Abdus Salam and J. Strathdee, Phys. Lett. 51B, 353 (1974), and, independently, by S. Ferrara and B. Zumino, Nucl. Phys. B79, 413 (1974).

[8] Our notational conventions are as follows. The Dirac matrices satisfy $\frac{1}{2}\{\gamma_\mu,\gamma_\nu\} = \eta_{\mu\nu} = \mathrm{diag}(+---)$. Ad-

joint spinors are defined by $\bar{\psi}=\psi^{\dagger}\gamma_0$. The matrices γ_0, $\gamma_0\gamma_\mu$, $\gamma_0\sigma_{\mu\nu}=\frac{1}{2}i\,\gamma_0[\gamma_\mu,\gamma_\nu]$, $\gamma_0 i\gamma_\mu\gamma_5$, and $\gamma_0\gamma_5=\gamma_1\gamma_2\gamma_3$ are Hermitian. The charge conjugate of ψ is defined by $\psi^c=C\bar{\psi}^T$, where $C^T=-C$ and $C^{-1}\gamma_\mu C=-\gamma_\mu^T$. By a Majorana spinor we mean $\psi^c=\psi$. It is useful to remember that the matrices $\gamma_\mu C$ and $\sigma_{\mu\nu}C$ are symmetric, while C, $\gamma_5 C$, and $i\gamma_\mu\gamma_5 C$ are antisymmetric. In particular, it follows that $\bar{\psi}\chi=\bar{\chi}\psi$, $\bar{\psi}\gamma_\mu\chi=-\bar{\chi}\gamma_\mu\psi$, $\bar{\psi}\sigma_{\mu\nu}\chi=-\bar{\chi}\sigma_{\mu\nu}\psi$, $\bar{\psi}i\gamma_\mu\gamma_5\chi=\bar{\chi}i\gamma_\mu\gamma_5\psi$, and $\bar{\psi}\gamma_5\chi=\bar{\chi}\gamma_5\psi$ if ψ and χ are anticommuting Majorana spinors.

[9]It may be remarked that the rank-4 "line element" $ds^2=(dx_\mu-\frac{1}{2}i\bar{\theta}\gamma_\mu d\theta)^2$ is an invariant.

[10]See, for example, J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Cambridge, Mass., 1955), Appendix A.2, p. 425.

[11]In certain cases it is possible to avoid this doubling and so define an intrinsic parity. The "vector" part, Φ_1, of the scalar superfield illustrates this [see Eqs. (3.10)–(3.13)].

[12]For renormalizability, V must be a polynomial of order 3. This Lagrangian, given by Wess and Zumino (Ref. 3), has been analyzed in detail by Iliopoulos and Zumino for renormalizability [J. Iliopoulos and B. Zumino, Nucl. Phys. B76, 310 (1974)]. See also Hung-Sheng Tsao, Phys. Lett. 53B, 381 (1974).

[13]Abdus Salam and J. Strathdee, Phys. Lett. 49B, 465 (1974), Footnote 2. A heuristic argument has been given by Iliopoulos and Zumino, suggesting that supersymmetry cannot be broken spontaneously (Ref. 12, Appendix 1).

[14]The compact notation becomes somewhat ambiguous in this formula. In particular, $\bar{V}_+\psi_+\equiv(C^{-1})^{\alpha\beta}V_{+\beta}\psi_{+\alpha}$.

[15]In order to avoid the Faddeev-Popov complications, it might be better to choose the noncovariant "axial" gauge $n_\mu A_\mu=0$ rather than $\partial_\mu A_\mu=0$.

[16]The same proposal was made independently by Ferrara and Zumino and by ourselves (Ref. 7). Another and quite different proposal is contained in R. Delbourgo, Abdus Salam, and J. Strathdee, Phys. Lett. 51B, 475 (1974).

[17]See Delbourgo, Salam, and Strathdee, Ref. 16.

Nuclear Physics B86 (1975) 142–152.North-Holland Publishing Company

FEYNMAN RULES FOR SUPERFIELDS

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and
Imperial College, London, England

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 16 September 1974

Abstract: The notion of functional differentiation with respect to superfields is defined and used to set up formal rules for computing Green functions and scattering amplitudes in supersymmetric models.

1. Introduction

A number of renormalizable Lagrangian models have been given recently which exhibit an underlying symmetry between fermions and bosons [1]. Although this so-called supersymmetry has yet to manifest itself in the real world, it is characterized by some unusual features which make it worthy of study [2].

Fundamentally, supersymmetry represents an extension of the Poincaré group [3] and could, indeed, be looked upon as a reflection of the structure of space-time. In this quasi-geometrical view, the extended Poincaré group can be made to act upon an 8-dimensional "extended spacetime" whose points are labelled by the pair (x_μ, θ_α), where x_μ denotes the usual minkowskian coordinate and θ_α is an *anticommuting* c-number Majorana spinor. The action of the new transformations on this space is given by

$$x_\mu \to x_\mu + \tfrac{1}{2} i \bar{\epsilon} \gamma_\mu \theta \,, \qquad \theta_\alpha \to \theta_\alpha + \epsilon_\alpha \,, \tag{1.1}$$

where the parameter ϵ_α is an anticommuting Majorana spinor. Our purpose here is to discuss the formulation of field theories on the extended space-time*.

*We shall be concerned with linear realizations only. One can avoid the extension of space-time by working with non-linear realizations wherein the supersymmetry is broken spontaneously. This approach was taken by Volkov and Akulov, ref. [3], and extended by Zumino, ref. [2].

A typical field, the scalar $\Phi(x,\theta)$, may be expanded in powers of θ. Owing to the anticommutation property, $\{\theta_\alpha,\theta_\beta\}=0$, such expansions must terminate. In fact only sixteen independent monomials can be made from the θ's. A convenient form [4*] of the expansion is given by

$$\Phi(x,\theta)=A(x)+\bar\theta\psi(x)$$
$$+\tfrac{1}{4}\bar\theta\theta F(x)+\tfrac{1}{4}\bar\theta\gamma_5\theta G(x)+\tfrac{1}{4}\bar\theta i\gamma_\nu\gamma_5\theta A_\nu(x)$$
$$+\tfrac{1}{4}\bar\theta\theta\,\bar\theta\chi(x)+\tfrac{1}{32}(\bar\theta\theta)^2 D(x)\,, \tag{1.2}$$

where the coefficients are ordinary fields: A, F, D are scalars, G is a pseudoscalar, A_ν is an axial vector, ψ and χ are Dirac spinors. One sees that half of the components are bosons and the other half fermions. The new dimensions of the extended space-time thus give rise to new degrees of freedom that are spin-like and finite in number.

It is possible to define the operation of differentiation with respect to the anticommuting coordinate θ and, in particular, the covariant derivative

$$D_\alpha\Phi=\frac{\partial\Phi}{\partial\bar\theta^\alpha}-\frac{1}{2}i(\gamma_\mu\theta)_\alpha\frac{\partial\Phi}{\partial x_\mu}\,. \tag{1.3}$$

This operator transforms as a spinor under the action of the Lorentz group and is invariant with respect to the "supertranslations" (1.1) whose action on Φ is given by

$$\delta\Phi=-\bar\epsilon^\alpha\left(\frac{\partial\Phi}{\partial\bar\theta^\alpha}+\frac{1}{2}i(\gamma_\mu\theta)_\alpha\frac{\partial\Phi}{\partial x_\mu}\right). \tag{1.4}$$

A number of properties of the operator D_α is derived in ref. [4]. The most important of these is the anticommutation rule

$$\{D_\alpha,D_\beta\}=-(\gamma_\mu C)_{\alpha\beta}\,i\frac{\partial}{\partial x_\mu}, \tag{1.5}$$

where C denotes the charge conjugation matrix.

The operator D_α is used to define the so-called chiral superfields $\Phi_\pm(x,\theta)$ by means of the covariant differential constraints

$$(1\mp i\gamma_5)D\,\Phi_\pm(x,\theta)=0\,. \tag{1.6}$$

These equations imply a number of relations among the coefficients in the expansion (1.2), *viz.*,

$$G_\pm=\pm iF_\pm\,,\qquad A_{\nu\pm}=\pm i\,\partial_\nu A_\pm\,,\qquad D_\pm=-\partial^2 A_\pm\,,\qquad x_\pm=-i\not\partial\psi_\pm\,,$$

where $(1\mp i\gamma_5)\psi_\pm=0$. It follows that the chiral superfields can be represented by the expansions:

*We follow the notation of this reference.

$$\Phi_{\pm}(x,\theta) = \exp\left(\mp \tfrac{1}{4}\bar{\theta}\gamma_\mu\gamma_5\theta\,\frac{\partial}{\partial x_\mu}\right)(A_{\pm}(x) + \bar{\theta}\psi_{\pm}(x) + \tfrac{1}{4}\bar{\theta}(1 \pm i\gamma_5)\theta\, F_{\pm}(x)) \,. \quad (1.7)$$

The set of chiral superfields is closed under multiplication,

$$\Phi_{+}(x,\theta)\,\Phi'_{+}(x,\theta) = \Phi''_{+}(x,\theta) \,.$$

This property, which is a direct consequence of the linear defining condition (1.6), is very useful in the construction of renormalizable Lagrangians, and the following sections will deal exclusively with chiral superfields.

It can be objected that the proposed transformation law (1.1) is not consistent in that to the coordinates x_μ, which are usually taken to be real numbers, are added the quantities $\frac{1}{2} i\,\bar{\epsilon}\gamma_\mu\theta$ which are *not* numbers. (Notice, for example, that $(\bar{\epsilon}\gamma_\mu\theta)^5 = 0$.) Clearly some generalized interpretation of the space-time coordinate x_μ is needed here. However, this is not a matter of any practical importance since one could easily suppress all reference to anticommuting c-numbers. Physical consequences of the supersymmetry could be extracted by requiring the scattering operator to commute with the "supercharge", S_α, whose action on the field components is defined by substituting

$$\delta\Phi = [\Phi, i\bar{\epsilon}^{\alpha} S_\alpha]$$

on the left-hand side of (1.4) and comparing coefficients. (For consistency, one must suppose that the spinorial quantities ϵ and θ anticommute amoung themselves as well as with the fermionic components ψ and χ.) The commutation rules satisfied by P_μ and $J_{\mu\nu}$, the generators of infinitesimal Poincaré transformations, must be supplemented by new rules involving S_α. These are:

$$[S_\alpha, P_\mu] = 0 \,, \qquad [S_\alpha, J_{\mu\nu}] = \tfrac{1}{2}\,(\sigma_{\mu\nu} S)_\alpha \,, \qquad \{S_\alpha, S_\beta\} = -(\gamma_\mu C)_{\alpha\beta} P_\mu \,. \quad (1.8)$$

The problem of constructing unitary representations of this system and deriving selection rules has been considered elsewhere [5]. In this paper we are attempting to show that adherence to superfields and anticommuting c-numbers does not appear to lead to any difficulties.

Up to now the concept of the superfield has played a subsidiary though useful role in guiding the construction of Lagrangian models. In effect these fields have been used to make action functionals which are manifestly supersymmetric. To proceed with any dynamical calculation, however, one was forced to relinquish the manifest symmetry and express the Lagrangian in terms of component Fermi and Bose fields of the usual kind. With the Lagrangian so expressed, one could follow the standard programme for computing Green functions and scattering amplitudes. This approach seems to us a retrograde one. It would be more satisfying if one could set up the Feynman rules and make calculations in the manifestly covariant notation. The purpose of this paper is to show that such a course can indeed be followed. We do not claim any computational advantage for the methods presented here. Rather,

we should like this formulation to be seen as a preliminary exploration of a methodology wherein supersymmetry is realized in a succint and obvious manner.

2. The functional derivative

For simplicity, we shall begin by considering functionals of a neutral scalar supermultiplet $\Phi_\pm(x,\theta)$. The chiral components of $\Phi_\pm$ are defined by the expansions

$$\Phi_\pm(x,\theta) = \exp\left[\mp\tfrac{1}{4}\,\bar\theta\partial\gamma_5\theta\right]\left(A_\pm(x)+\bar\theta\,\tfrac{1}{2}(1\pm i\gamma_5)\,\psi(x)+\tfrac{1}{4}\bar\theta(1\pm i\gamma_5)\theta\,F_\pm(x)\right), \tag{2.1}$$

where $\psi(x)$ is a Majorana spinor and $A_\pm, F_\pm$ are complex boson fields subject to the reality conditions $A_- = A_+^*, F_- = F_+^*$. Under space reflections the fields A_+ and F_+ are carried into A_- and F_-, respectively.

Let Γ be a functional of Φ_+ and Φ_-. Corresponding to the infinitesimal variations $\delta\Phi_\pm$ the variation in Γ may be written†

$$\delta\Gamma = \int dx\,(-\tfrac{1}{2}\bar D D)\left(\delta\Phi_+(x,\theta)\,\frac{\delta\Gamma}{\delta\Phi_+(x,\theta)} + \delta\Phi_-(x,\theta)\,\frac{\delta\Gamma}{\delta\Phi_-(x,\theta)}\right). \tag{2.2}$$

This formula serves to define the derivatives of Γ with respect to Φ_+ and Φ_-. These derivatives, $\delta\Gamma/\delta\Phi_\pm$, are themselves chiral superfields and their components may be expressed in terms of the derivatives of Γ with respect to the component fields $A_\pm(x)$, etc. It is natural to define these latter derivatives by the linear form,

$$\delta\Gamma = \int dx\left(\delta A_+(x)\,\frac{\delta\Gamma}{\delta A_+(x)} + \delta A_-(x)\,\frac{\delta\Gamma}{\delta A_-(x)} + \delta\bar\psi(x)\,\frac{\delta\Gamma}{\delta\bar\psi(x)}\right.$$

$$\left. + \delta F_+(x)\,\frac{\delta\Gamma}{\delta F_+(x)} + \delta F_-(x)\,\frac{\delta\Gamma}{\delta F_-(x)}\right). \tag{2.3}$$

Comparing the expressions (2.2) and (2.3) one finds

$$\frac{\delta\Gamma}{\delta\Phi_\pm(x)} = \exp\left[\mp\tfrac{1}{4}\,\bar\theta\partial\gamma_5\theta\right]\left(\frac{\delta\Gamma}{\delta F_\pm(x)} - \bar\theta\,\tfrac{1}{2}(1\pm i\gamma_5)\,\frac{\delta\Gamma}{\delta\bar\psi(x)}\right.$$

$$\left. + \tfrac{1}{4}\,\bar\theta\,(1\pm i\gamma_5)\,\theta\,\frac{\delta\Gamma}{\delta A_\pm(x)}\right). \tag{2.4}$$

This formula serves quite generally for translating ordinary functional derivatives into the superfield form. A particular application of it – with Γ set equal to

†This integral defines an invariant bilinear form. The forms

$$\int dx\,(-\tfrac{1}{2}\bar D D)\,(\Phi_\pm^{(1)}\,\Phi_\pm^{(2)}) = \int dx\,(A_\pm^{(1)}\,F_\pm^{(2)} + F_\pm^{(1)}\,A_\pm^{(2)} - \bar\psi_\pm^{(1)c}\,\psi_\pm^{(2)})$$

are invariant under supertransformations and proper Lorentz transformations.

$\Phi_\pm(x',\theta')$ – serves to define what might be called the "super delta-function". Thus, one finds

$$\frac{\delta\Phi_\pm(x',\theta')}{\delta\Phi_\pm(x,\theta)} = \exp\left[\mp\tfrac{1}{4}\,\bar\theta\partial\gamma_5\theta \mp \tfrac{1}{4}\,\bar\theta'\partial'\gamma_5\theta'\right]\tfrac{1}{4}(\bar\theta-\bar\theta')(1\pm i\gamma_5)(\theta-\theta)\,\delta(x-x')$$

$$\equiv \delta_\pm(x,\theta;x'\theta')\,, \tag{2.5}$$

and, on the other hand,

$$\frac{\delta\Phi_\mp(x',\theta')}{\delta\Phi_\pm(x,\theta)} = 0\,. \tag{2.6}$$

The functions $\delta_\pm(x,\theta;x'\theta')$ enjoy the following properties:

(a) symmetry under the interchange $(x,\theta) \leftrightarrow (x',\theta')$;

(b) Poincaré invariance;

(c) supersymmetry, *viz.*,

$$\delta_\pm(x+\tfrac{1}{2}\,i\,\epsilon\gamma\theta,\theta+\epsilon;x'+\tfrac{1}{2}\,i\,\epsilon\gamma\theta',\theta'+\epsilon) = \delta_\pm(x,\theta;x',\theta')\,; \tag{2.7}$$

(d) chirality

$$(1\mp i\gamma_5)\,D\,\delta_\pm(x,\theta;x',\theta') = 0\,; \tag{2.8}$$

(e) the integral identities

$$\int dx\,(-\tfrac{1}{2}\,\bar D D)\,(\psi_\pm(x,\theta)\,\delta_\pm(x,\theta;x',\theta')) = \psi_\pm(x',\theta') \tag{2.9}$$

for arbitrary chiral superfields $\psi_\pm$ (this means that the functions $\delta_\pm$ serve as identity distributions).

It would appear that, although we do not have a measure of the usual sort on the space of x and θ, it is possible to treat the operation

$$\int dx\,(-\tfrac{1}{2}\,\bar D D)$$

as a kind of generalized integration. Pursuing the analogy, we may define the functional Taylor expansion,

$$\Gamma(\Phi_+,\Phi_-) = \sum_{n,m}\frac{1}{n!m!}\prod_1^{n+m}\int dx_j(-\tfrac{1}{2}\,\bar D_j D_j)[K_{nm}(x_1\theta_1,\ldots,x_{n+m}\,\theta_{n+m})$$

$$\times\,\Phi_+(x_1\theta_1)\ldots\Phi_+(x_n\theta_n)\,\Phi_-(x_{n+1}\,\theta_{n+1})\ldots\Phi_-(x_{n+m}\,\theta_{n+m})]\,, \tag{2.10}$$

where the coefficient distributions K_{nm} satisfy the chiral conditions

$$(1-i\gamma_5)\,D_j K_{nm}(1,\ldots,n+m) = 0, \qquad \text{for} \qquad j=1,\ldots,n\,,$$

$$(1+i\gamma_5)\,D_j K_{nm}(1,\ldots,n+m) = 0 \qquad \text{for} \qquad j=n+1,\ldots,n+m\,.$$

A typical example of a functional defined over the superfields $\Phi_\pm$ is the classical (local) action

$$S(\Phi_+, \Phi_-) = \int dx\, [\tfrac{1}{8}(\bar{D}D)^2(\Phi_-\Phi_+) - \tfrac{1}{2}\bar{D}D\,(V(\Phi_+) + V(\Phi_-))]\,, \tag{2.11}$$

where V is a polynomial*. To evaluate the derivatives of S it is necessary to make use of the operator identities [4]

$$\tfrac{1}{2}(\bar{D}D)^2 = \bar{D}\tfrac{1}{2}(1-i\gamma_5)D\,\bar{D}\tfrac{1}{2}(1+i\gamma_5)D + \bar{D}\not{\partial}\gamma_5 D$$

$$= \bar{D}\tfrac{1}{2}(1+i\gamma_5)D\,\bar{D}\tfrac{1}{2}(1-i\gamma_5)D - \bar{D}\not{\partial}\gamma_5 D\,. \tag{2.12}$$

Make an infinitesimal variation in (2.11),

$$\delta S = \int dx\, [\tfrac{1}{8}(\bar{D}D)^2(\delta\Phi_-\Phi_+ + \delta\Phi_+\Phi_-) - \tfrac{1}{2}\bar{D}D\,(\delta\Phi_+ V'(\Phi_+) + \delta\Phi_- V'(\Phi_-))]$$

$$= \int dx\, [\tfrac{1}{4}\bar{D}\tfrac{1}{2}(1-i\gamma_5)D\bar{D}\tfrac{1}{2}(1+i\gamma_5)D\,(\delta\Phi_-\Phi_+)$$

$$+ \tfrac{1}{4}\bar{D}\tfrac{1}{2}(1+i\gamma_5)D\bar{D}\tfrac{1}{2}(1-i\gamma_5)D\,(\delta\Phi_+\Phi_-)$$

$$+ \tfrac{1}{4}\bar{D}\not{\partial}\gamma_5 D\,(\delta\Phi_-\Phi_+ - \delta\Phi_+\Phi_-)$$

$$- \tfrac{1}{2}\bar{D}D\,(\delta\Phi_- V'(\Phi_-) + \delta\Phi_+ V'(\Phi_+))]\,.$$

Making use of the chiral properties of $\Phi_\pm$ and $\delta\Phi_\pm$ one can reduce this immediately to the form

$$\delta S = \int dx\,(-\tfrac{1}{2}\bar{D}D)\,[\delta\Phi_+(-\tfrac{1}{2}\bar{D}D\,\Phi_- + V'(\Phi_+)) - \delta\Phi_-(-\tfrac{1}{2}\bar{D}D\,\Phi_+ + V'(\Phi_-))]$$

$$+ \oint d\Sigma_\mu\, \tfrac{1}{4}\bar{D}\gamma_\mu\gamma_5 D\,(\Phi_+\delta\Phi_- - \Phi_-\delta\Phi_+)\,. \tag{2.13}$$

Discarding the surface term** and comparing the remainder with (2.2), one obtains the functional derivatives

$$\frac{\delta S}{\delta\Phi_\pm} = -\tfrac{1}{2}\bar{D}D\,\Phi_\mp + V'(\Phi_\pm)\,. \tag{2.14}$$

Setting these equal to zero gives the classical equations of motion in superfield form. In sect. 3 the problem of quantizing this system will be considered.

The functional derivatives with respect to chiral superfields are now seen to be satisfactorily defined by (2.1). Derivatives with respect to non-chiral superfields, however, must be independently defined. Such non-chiral fields are needed to represent the gauge dependent potentials in models which carry a local symmetry [4].

*One may ask whether the action functional (2.11) can be expressed in the multilocal form (2.10). The answer is, indeed, yes. The coefficient K_{11}, in particular, is given by

$$K_{11}(1,2) = -\tfrac{1}{4}(\bar{D}D)_1\,\delta_-(1,2) - \tfrac{1}{4}(\bar{D}D)_2\,\delta_+(1,2)\,.$$

**The surface term is of course important for the derivation of conserved currents but we shall not consider them here.

A suitable definition for the derivative with respect to the non-chiral field $\psi(x,\theta)$ is given by

$$\delta\Gamma = \int \mathrm{d}x\, \tfrac{1}{2}(\bar{D}D)^2\,(\delta\psi(x,\theta)\,\delta\Gamma/\delta\psi(x,\theta))\,. \tag{2.15}$$

With the component structure of ψ defined by the expansion (1.2) and derivatives with respect to these components defined by:

$$\delta\Gamma = \int \mathrm{d}x \left(\delta A \frac{\delta\Gamma}{\delta A} + \delta\bar{\psi}\frac{\delta\Gamma}{\delta\bar{\psi}} + \delta F \frac{\delta\Gamma}{\delta F} + \delta G \frac{\delta\Gamma}{\delta G} + \delta A_\nu \frac{\delta\Gamma}{\delta A_\nu} + \delta\bar{\chi}\frac{\delta\Gamma}{\delta\bar{\chi}} + \delta D \frac{\delta\Gamma}{\delta D}\right), \tag{2.16}$$

one finds the following component structure for the derivative with respect to $\psi(x,\theta)$:

$$\begin{aligned}\frac{\delta\Gamma}{\delta\bar{\psi}(x,\theta)} &= \frac{\delta\Gamma}{\delta D(x)} - \tfrac{1}{2}\bar{\theta}\,\frac{\delta\Gamma}{\delta\bar{\chi}(x)} \\ &+ \tfrac{1}{8}\bar{\theta}\theta\,\frac{\delta\Gamma}{\delta F(x)} + \tfrac{1}{8}\bar{\theta}\gamma_5\theta\,\frac{\delta\Gamma}{\delta G(x)} + \tfrac{1}{8}\bar{\theta}i\gamma_\nu\gamma_5\theta\,\frac{\delta\Gamma}{\delta A_\nu(x)} \\ &- \tfrac{1}{8}\bar{\theta}\theta\,\bar{\theta}\,\frac{\delta\Gamma}{\delta\bar{\psi}(x)} + \tfrac{1}{32}(\bar{\theta}\theta)^2\,\frac{\delta\Gamma}{\delta A(x)}\,.\end{aligned} \tag{2.17}$$

The corresponding super delta-function is given by

$$\frac{\delta\psi(x,\theta)}{\delta\psi(x',\theta')} = \delta(x,\theta;x',\theta') = \tfrac{1}{32}\,((\bar{\theta}-\bar{\theta}')\,(\theta-\theta'))^2\,\delta(x-x')\,. \tag{2.18}$$

This invariant function acts as the identity distribution,

$$\int \mathrm{d}x\, \tfrac{1}{2}(\bar{D}D)^2\,(\psi(x,\theta)\,\delta(x,\theta;x',\theta')) = \psi(x',\theta')\,. \tag{2.19}$$

3. Covariant quantization

The formal derivation of Feynman rules proceeds very much as in any ordinary quantized field theory. To illustrate this we consider the case of a neutral scalar supermultiplet whose dynamics at the classical level is governed by the action functional (2.11). A straightforward method to obtain the Feynman rules is to perturb the system by means of an external current distribution, $J_\pm(x,\theta)$,

$$S \to S + \int \mathrm{d}x\,(-\tfrac{1}{2}\bar{D}D)\,[J_+\Phi_+ + J_-\Phi_-]\,,$$

and represent the corresponding vacuum transition amplitude by a path integral,

$$\exp\frac{i}{\hbar}Z(J_+,J_-) = \int(\mathrm{d}\Phi_+\,\mathrm{d}\Phi_-)\exp\frac{i}{\hbar}\,[S(\Phi_+,\Phi_-) + \int \mathrm{d}x\,(-\tfrac{1}{2}\bar{D}D)\,(J_+\Phi_+ + J_-\Phi_-)]\,. \tag{3.1}$$

One proceeds by separating the classical action into a free (bilinear) and an interaction piece,

$$V(\Phi_\pm) = \tfrac{1}{2} M \Phi_\pm^2 + V_{\text{int}}(\Phi_\pm)\,, \tag{3.2}$$

where V_{int} contains cubic and higher order terms*. In (3.1) make the formal replacement

$$V_{\text{int}}(\Phi_\pm) \to V_{\text{int}}\left(\frac{\hbar}{i}\frac{\delta}{\delta J_\pm}\right),$$

so as to represent the vacuum amplitude in the form

$$\exp\frac{i}{\hbar}Z(J_+,J_-) = \left[\exp\frac{i}{\hbar}\int dx\,(-\tfrac{1}{2}\bar{D}D)\left(V_{\text{int}}\left(\frac{\hbar}{i}\frac{\delta}{\delta J_+}\right) + V_{\text{int}}\left(\frac{h}{i}\frac{\delta}{\delta J_+}\right)\right)\right]$$

$$\times \exp\frac{i}{\hbar}Z_0(J_+,J_-)\,, \tag{3.3}$$

where Z_0 is represented by a Gaussian path-integral,

$$\exp\frac{i}{\hbar}Z_0(J_+,J_-) = \int (d\Phi_+ d\Phi_-)\exp\frac{i}{\hbar}\int dx\,[\tfrac{1}{8}(\bar{D}D)^2(\Phi_+\Phi_-)$$

$$-\tfrac{1}{2}\bar{D}D\,(\tfrac{1}{2}M(\Phi_+^2+\Phi_-^2) + (J_+\Phi_+ + J_-\Phi_-))]\,. \tag{3.4}$$

Being Gaussian, this integral can be evaluated by the usual trick of translating the integration variables. An equivalent method is to obtain functional differential eauations for Z_0 from the identities

$$0 = \int (d\Phi_+ d\Phi_-)\frac{\delta}{\delta\Phi_\pm(x',\theta')}\exp\frac{i}{\hbar}\int dx\,[\tfrac{1}{8}(\bar{D}D)^2(\Phi_+\Phi_-)$$

$$-\tfrac{1}{2}\bar{D}D\,(\tfrac{1}{2}M(\Phi_+^2+\Phi_-^2) + (J_+\Phi_+ + J_-\Phi_-))]$$

$$= \int (d\Phi_+ d\Phi_-)\frac{i}{\hbar}[-\tfrac{1}{2}\bar{D}D\Phi_\mp + M\Phi_\pm + J_\pm]\exp\frac{i}{\hbar}\int dx\,[''']$$

$$= \frac{i}{\hbar}\left[-\tfrac{1}{2}\bar{D}D\frac{\hbar}{i}\frac{\delta}{\delta J_\mp} + M\frac{\hbar}{i}\frac{\delta}{\delta J_\pm} + J_\pm\right]\exp\frac{i}{\hbar}Z_0$$

$$= \frac{i}{\hbar}\left(\exp\frac{i}{\hbar}Z_0\right)\left[-\tfrac{1}{2}\bar{D}D\frac{\delta Z_0}{\delta J_\mp} + M\frac{\delta Z_0}{\delta J_\pm} + J_\pm\right].$$

The linear inhomogeneous equations

$$-\tfrac{1}{2}\bar{D}D\frac{\delta Z_0}{\delta J_\mp} + M\frac{\delta Z_0}{\delta J_\pm} = -J_\pm\,, \tag{3.5}$$

*In general the expansion of S must be taken about a stationary point. For simplicity, we are here assuming that $\Phi_\pm = 0$ is such a point.

are easily solved (with the help of the identity, $(\bar{D}D)^2 = -4\,\partial^2$, on chiral fields [4]) to give

$$\frac{\delta Z_0}{\delta J_\pm} = \frac{1}{\partial^2 + M^2}\left(-\tfrac{1}{2}\bar{D}D\,J_+ - M J_\pm\right). \tag{3.6}$$

These equations in turn can be integrated to give

$$Z_0(J_+, J_-) = \int \mathrm{d}x \left[\tfrac{1}{8}(\bar{D}D)^2\left(J_+ \frac{1}{\partial^2 + M^2} J_-\right)\right.$$

$$\left. -\tfrac{1}{2}\bar{D}D\left(J_+ \frac{-\frac{1}{2}M}{\partial^2 + M^2} J_+ + J_- \frac{-\frac{1}{2}M}{\partial^2 + M^2} J_-\right)\right]. \tag{3.7}$$

The perturbation development of the connected vacuum amplitude $(i/\hbar)Z$ is obtained from (3.3) by expanding in powers of V_{int} and substituting the expression (3.7) for Z_0. The bare propagators are of course the secondorder functional derivatives of Z_0. Their explicit forms are given by

$$\frac{i}{\hbar}\langle \mathrm{T}\Phi_\pm(x,\theta)\,\Phi_\pm(x',\theta')\rangle_0 = \frac{-M}{\partial^2 + M^2}\delta_\pm(x,\theta;x',\theta'),$$

$$\frac{i}{\hbar}\langle \mathrm{T}\Phi_\pm(x,\theta)\,\Phi_\mp(x',\theta')\rangle_0 = \frac{1}{\partial^2 + M^2}\left(-\tfrac{1}{2}\bar{D}D\right)\delta_\mp(x,\theta;x'\theta'), \tag{3.8}$$

where the delta-functions $\delta_\pm$ are defined in (2.5). In terms of the causal function

$$\Delta_{\mathrm{F}}(x - x'; M) = \int \frac{\mathrm{d}_4 k}{(2\pi)^4} \frac{i\hbar}{k^2 - M^2 + i\epsilon} \exp\left[-ik\,(x - x')\right],$$

the propagators (3.8) can be written,

$$\langle \mathrm{T}\Phi_\pm(x,\theta)\,\Phi_\pm(x',\theta')\rangle_0$$

$$= -M\left[\exp\left(\mp\tfrac{1}{4}\right)(\bar{\theta}\not{\partial}\gamma_5\theta + \bar{\theta}'\not{\partial}'\gamma_5\theta')\right]\tfrac{1}{4}(\bar{\theta} \mp \bar{\theta}')(1 \pm i\gamma_5)(\theta - \theta')\,\Delta_{\mathrm{F}},$$

$$\langle \mathrm{T}\Phi_\pm(x,\theta)\,\Phi_\mp(x',\theta')\rangle_0$$

$$= \left[\exp\left(\mp\tfrac{1}{4}\right)(\bar{\theta}\not{\partial}\gamma_5\theta - \bar{\theta}'\not{\partial}'\gamma_5\theta')\right]\left[\exp\tfrac{1}{2}i\,\bar{\theta}\,(1 \pm i\gamma_5)\not{\partial}\theta'\right]\Delta_{\mathrm{F}}. \tag{3.9}$$

These expressions are to be associated with the internal lines of any graph. Vertices, on the other hand, are associated with the expression

$$\frac{i}{\hbar} g_n \int \mathrm{d}x\,(-\tfrac{1}{2}\bar{D}D), \tag{3.10}$$

where g_n is a coupling constant from the expansion*

*If the model is to be renormalizable, then only g_3 can differ from zero.

$$V_{\text{int}}(\Phi_{\pm}) = \frac{g_3}{3!}\Phi_{\pm}^3 + \frac{g_4}{4!}\Phi_{\pm}^4 + \dots \quad . \tag{3.11}$$

The integration over x at each vertex - which effects the conservation of four-momentum – gives rise to some simplifications. Many of the exponentiated factors in (3.9) can be seen to cancel so that the internal lines are associated with the "effective" propagators

$$\langle T\Phi_{\pm}(x,\theta)\,\Phi_{\pm}(x',\theta')\rangle_{\text{eff}} = -\tfrac{1}{4}M(\bar{\theta}-\bar{\theta}')(1\pm i\gamma_5)(\theta-\theta')\,\Delta_{\text{F}}\,,$$

$$\langle T\Phi_{\pm}(x,\theta)\,\Phi_{\mp}(x',\theta')\rangle_{\text{eff}} = [\exp\tfrac{1}{2}i\,\bar{\theta}(1\pm i\gamma_5)\not{\partial}\theta']\,\Delta_{\text{F}}\,. \tag{3.12}$$

External lines should be associated with the wave functions

$$\varphi_{\pm}^{\text{ext}}(x,\theta) = A_{\pm}^{\text{ext}}(x) + \bar{\theta}\,\tfrac{1}{2}(1\pm i\gamma_5)\,\psi^{\text{ext}}(x) + \tfrac{1}{4}\bar{\theta}(1\pm i\gamma_5)\theta\,F_{\pm}^{\text{ext}}(x)\,. \tag{3.13}$$

The momentum space Feynman rules may be summarized as follows:

(i) Draw diagrams in the usual way with vertices corresponding to the various types of monomial $(g_n/n!)\,\Phi_+^n$ and $(g_n/n!)\,\Phi_-^n$ in the interaction Lagrangian.

(ii) With each n-leg vertex associate the factor

$$(2\pi)^4\,\delta(p_1+\dots+p_n)\frac{i}{\hbar}\,g_n\,.$$

(iii) With each line joining a pair of vertices of the same chirality (±) associate the effective propagator

$$(\hbar/i)\,(\tfrac{1}{2}M)\,(\bar{\theta}_1-\bar{\theta}_2)\,\tfrac{1}{2}(1\pm i\gamma_5)(\theta_1-\theta_2)\,1/(-p^2+M^2-i\epsilon)\,.$$

(iv) With each line joining a pair of vertices with opposite chirality associate the effective propagator

$$(\hbar/i)\exp(\theta_1\,\tfrac{1}{2}(1+i\gamma_5)\,p\,\theta_2)\,1/(-p^2+M^2-i\epsilon)\,,$$

where the four-momentum is directed from the + type vertex (1) to the – type vertex (2).

(v) With each external line associate the wave function

$$\varphi_{\pm}^{\text{ext}} = A_{\pm}^{\text{ext}}(p) + \bar{\theta}\,\tfrac{1}{2}(1\pm i\gamma_5)\,\psi^{\text{ext}}(p) + \tfrac{1}{2}\bar{\theta}\,\tfrac{1}{2}(1\pm i\gamma_5)\,\theta\,F_{\pm}^{\text{ext}}(p)\,.$$

(vi) At each vertex apply the differential operator

$$-\tfrac{1}{2}\bar{D}D = -\tfrac{1}{2}(C^{-1})^{\beta\alpha}\,\frac{\partial}{\partial\bar{\theta}^{\alpha}}\,\frac{\partial}{\partial\bar{\theta}^{\beta}}\,,$$

i.e. extract the coefficient of the term

$$\prod_{\text{vertices}}\tfrac{1}{4}\bar{\theta}(1\pm i\gamma_5)\theta\,.$$

(vii) Integrate over the momenta,

$$\prod_{\text{lines}}\int\frac{\mathrm{d}p}{(2\pi)^4}\,.$$

To pick out a given kind of process, one would of course specialize the external wave functions. For example, an external on-shell 0^+ particle would be represented by the wave functions

$$\varphi_{\pm}^{\text{ext}}(x,\theta) = \frac{1}{(2\pi)^3} \exp\left[-ikx\right](1 - \tfrac{1}{4} M\,\bar{\theta}\,(1 \pm i\gamma_5)\,\theta)\,.$$

The efficacy of this kind of approach has been demonstrated in the work of Capper and Delbourgo [6] who show that the divergence softening characteristic of supersymmetric models can be understood in terms of the properties of superfield graphs. In particular, Capper and Leibbrandt [7] have developed a formula for the superficial degree of divergence of such graphs.

References

[1] J. Wess and B. Zumino, Phys. Letters 49B (1974) 52; Nucl. Phys. B78 (1974) 1;
Abdus Salam and J. Strathdee, ICTP, Trieste, preprint IC/74/36, Phys. Letters, to appear;
S. Ferrara and B. Zumino, Nucl. Phys. B79 (1974) 413;
R. Delbourgo, Abdus Salam and J. Strathdee, ICTP, Trieste, preprint IC/74/45, Phys. Letters, to appear;
P. Fayet and J. Iliopoulos, Lab. de Physique Théorique et Hautes Energies, Orsay, preprint LPTHE 74/26 (PTENS 74/3).

[2] B. Zumino, Invited talk at the 17th Int. Conf. on high-energy Physics, London (1974), CERN preprint TH 1901 (1974).

[3] D.V. Volkov and V.P. Akulov, Phys. Letters 46B (1973) 109.

[4] Abdus Salam and J. Strathdee, ICTP, Trieste, preprint IC/74/42, Phys. Rev., to be published.

[5] Abdus Salam and J. Strathdee, Nucl. Phys. B80 (1974) 499.

[6] D. Capper, ICTP, Trieste, preprint IC/74/66;
R. Delbourgo, Imperial College, London, preprint ICTP/73/44.

[7] D. Capper and G. Leibbrandt, ICTP, Trieste, preprint IC/74/79.

Volume 61B, number 4 PHYSICS LETTERS 12 April 1976

BLACK HOLES AS SOLITONS

A. SALAM
International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 2 February 1976

We remark that exact classical Schwarzschild-like solutions to Einstein's (and possibly f gravity) equations provide examples of realistic solitons.

Under the broadest definition, any non-trivial solution to a system of classical non-linear equations, which is confined to a finite region of space and which carries a finite energy, may be considered a soliton. The problem is to discover to what extent such classical objects can approximate to the quantum systems encountered in particle physics. Are they stable? What conserved quantities can be associated with them? How do they interact with "ordinary" particles described by quantized fields?

It is the purpose of this note to point out that a whole class of solitons is apparently available among the known exact classical Schwarzschild-like solutions to Einstein's equations. These solutions ara stationary, possess finite energy and can be localized. The stability of some of these solutions depends on rather general considerations but, in any case, the association of a finite and conserved energy with asymptotically flat (i.e. localized) solutions follows from the existence of the so-called energy-momentum pseudo-tensor.

For solutions to Einstein's equations, the conservation of energy and momentum results from appropriate symmetries in the equations of motion. This is unlike the conservation of magnetic charge in 't Hooft's theory [1], where electromagnetism is part of a spontaneously broken non-abelian gauge system. In such theories the conservation is not dependent on a symmetry of the Lagrangian – and, moreover, magnetic charge is confined to sites of topological significance. There has been some controversy as to whether such topological significance is the hallmark of a soliton, (see, for example, Lee [2] for the contrary viewpoint).

Reprinted with permission from *Phys. Lett.* **61B** (1976) pp. 375–376
© 1976 Elsevier Science Publishers B.V.

It is worth remarking in this context that, although classically stable, the gravitational solitons may be able to decay through vacuum effects. Thus, Hawking [3] has shown that the Schwarzschild solution radiates like a black body whose temperature is inversely proportional to its mass. Conceivably, vacuum effects could be relevant for non-gravitational solitons; for example, stability may be lost for solitons carrying topological conserved numbers ('t Hooft's monopole) through a phase transition induced by a high temperature or strong external electromagnetic environment.

Gravitational forces should not be significant in particle physics at foreseeable energies*. However, the mathematically analogous f gravity [4] associated with strongly interacting 2^+ mesons could be relevant. Indeed, it was suggested some years ago that f gravity black holes – now to be called solitons – might represent hadrons [4, 5]. For such black holes the "Schwarzschild radius" of an f gravity soliton would be comparable to its Compton wavelength. An interesting possibility is that the well-known property of the black hole, whereby nothing can escape from it, could have implications for the confinement problem. The colour selectivity for saturation of strong forces (i.e. the colour singlet character of ordinary matter) could result from an $SL(6,C)'$ structure for f gravitons [6], where the relevant internal symmetry is $SU(3)'$ of colour.

* But note that masses up to the Planck value $\sim 10^{19}$ GeV are already being contemplated for unified models of weak and electromagnetic interactions.

Volume 61B, number 4 PHYSICS LETTERS 12 April 1976

At present there are no exact solutions known to the two-tensor f-g generally covariant equations and no solutions known even in the flat space approximation $g_{\mu\nu} = \eta_{\mu\nu}$. However, if the mass of the f meson is discarded then the equations for $f_{\mu\nu}$ become identical in this approximation with Einstein's equations for $g_{\mu\nu}$. We conjecture that it may be possible to restore the effects of the mass term by means of the replacement $1/r \to e^{-mr}/r$ in the Schwarzschild solution – at least in some favourable co-ordinate system.

The most interesting solution of the pure Einstein–Maxwell equations is the charged-Kerr–Newman rotating system [7] described by the gravitational and electromagnetic fields,

$$g_{\mu\nu}(x)\,dx^\mu dx^\nu = -(\Delta/\rho^2)(dt - a\sin^2\theta\,d\phi)^2 + (\sin^2\theta/\rho^2)((r^2+a^2)\,d\phi - a\,dt)^2 + (\rho^2/\Delta)\,dr^2 + \rho^2\,d\theta\,,$$

$$F_{\mu\nu}(x)\,dx^\mu \times dx^\nu = Q\rho^{-4}(r^2 - a^2\cos^2\theta)\,dr \times (dt - a\sin^2\theta\,d\phi) + 2Q\rho^{-4}ar\cos\theta\sin\theta\,d\theta \times ((r^2+a^2)\,d\phi - a\,dt)\,,$$

where $\Delta = r^2 - 2Mr + a^2 + Q^2$ and $\rho^2 = r^2 + a^2\cos^2\theta$. (These are the Boyer–Lindquist [8] or generalized Schwarzschild co-ordinates.) The energy and angular momentum carried by this soliton can be deduced from the asymptotic form of the metric fields. Thus, using the definition of the energy-momentum pseudotensor given by Landau and Lifshitz [9],

$$\Theta^{\nu\kappa}_{\text{total}} = \frac{\partial}{\partial x^\lambda} h^{\nu\kappa\lambda} = \frac{1}{16\pi}\frac{\partial^2}{\partial x^\lambda \partial x^\mu}[\sqrt{-g}(g^{\nu\kappa}g^{\lambda\mu} - g^{\nu\lambda}g^{\kappa\mu})],$$

the space integrals for P^μ and $M^{\mu\nu}$ can be reduced to two-dimensional surface integrals over a sphere of infinite radius. One finds a total energy equal to M and angular momentum equal to Ma. The linear momentum vanishes. In a similar fashion one deduces from the asymptotic form of $F_{\mu\nu}$ that the total charge is Q and magnetic moment Qa. (Notice that the gyromagnetic ratio takes the Dirac value, Q/M.)

In addition to these, there are other known classical solutions; for example Curzon's solution [10] of Einstein's equations:

$$ds^2 = \exp(-2m/\rho)(dt)^2 - \exp(2m/\rho)(\exp(-m^2r^2/2\rho^4)(dr^2+dz^2) + d\phi^2),$$
$$\rho^2 = r^2 + z^2\,,$$

and Singh's solution [11] of a scalar tensor theory:

$$ds^2 = \left(1 - \frac{2m}{r}\right)^{b/\sqrt{b^2-3/4}} dt^2 - \left(1 - \frac{2m}{r}\right)^{-b/\sqrt{b^2-3/4}} \times \left[dr^2 + \left(1 - \frac{2m}{r}\right) r^2 (d\theta^2 + \sin\theta^2\,d\phi^2)\right],$$
$$\Phi = \left(1 - \frac{2m}{r}\right)^{1/2\sqrt{b^2-3/4}},$$

this latter is related to Schwarzschild's solution. With these four-dimensional soliton solutions available, we feel that the recent work on particle production near black holes affords a particularly favourable testing ground for ideas in soliton physics.

We thank Dr. M.N. Mahanta for telling us of Singh's solution, and Drs. C.J. Isham and M.J. Duff for discussions.

References

[1] G. 't Hooft, Nucl. Phys. B79 (1974) 276.
[2] T.D. Lee, Columbia preprint CO-2271-60 (1975).
[3] S. Hawking, Nature 248 (1974) 30.
[4] C.J. Isham, A. Salam and J. Strathdee, Phys. Rev. D3 (1971) 867.
[5] A. Salam, Non-polynomial Lagrangians, renormalization and gravity (Gordon and Breach, London 1971), Vol. I, p. 31.
[6] C.J. Isham, A. Salam and J. Strathdee, Phys. Rev. D8 (1973) 2600.
[7] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (W.H. Freeman, San Francisco 1973).
[8] R.H. Boyer, W. Lindquist, J. Math. Phys. 8 (1967) 265.
[9] L. Landau and E. Lifshitz, The classical theory of fields (Addison–Wesley, Cambridge 1951).
[10] H.E.J. Curzon, Proc. London Math. Soc. 23 (1924) 477.
[11] T. Singh, J. Math. Phys. 16 (1975) 2109.

Reprinted from ANNALS OF PHYSICS
Vol. 141, No. 2, July 1982
All Rights Reserved by Academic Press, New York and London

On Kaluza–Klein Theory

ABDUS SALAM

International Centre for Theoretical Physics, Trieste, Italy, and Imperial College, London, England

AND

J. STRATHDEE

International Centre for Theoretical Physics, Trieste, Italy

Received December 8, 1981

Assuming the compactification of $4+K$-dimensional space-time implied in Kaluza–Klein-type theories, we consider the case in which the internal manifold is a quotient space, G/H. We develop normal mode expansions on the internal manifold and show that the conventional gravitational plus Yang–Mills theory (realizing local G symmetry) is obtained in the leading approximation. The higher terms in the expansions give rise to field theories of massive particles. In particular, for the original Kaluza–Klein $4+1$-dimensional theory, the higher excitations describe massive, charged, purely spin-2 particles. These belong to infinite dimensional representations of an $O(1,2)$.

I. INTRODUCTION

The motivation for studying generally covariant field theories in space-time of more than four dimensions is to obtain a geometrical interpretation of internal quantum numbers such as electric charge, i.e., to place them in the same context as energy and momentum [1, 2]. The latter observables are associated with translational symmetry in 4-dimensional Minkowski space; the internal observables would be associated with symmetry motions in the extra dimensions.

In theories of the Kaluza–Klein type one starts with the hypothesis that spacetime has $4+K$ dimensions. One assumes general covariance and adopts the $4+K$-dimensional curvature scalar as the Lagrangian. Next, one supposes that because of some dynamical mechanism, the ground state of this system is partially compactified [3], $M^4 \times B^K$ rather than M^{4+K}, where M^4 denotes 4-dimensional Minkowski space and B^K is a compact K-dimensional space [4]. The size of B^K must be sufficiently small to render it unresolvable at the currently available energies. For example, with the size of $B^K < 10^{-17}$ cm, it would be invisible to probes of energy $<10^3$ GeV, this being an upper limit.

Models have been constructed which illustrate this phenomenon of spontaneous

Copyright © 1982 by Academic Press, Inc.

compactification [5, 6]. In general, it is necessary for the system to include non-geometrical (matter) fields, coupled to the metric field, for this effect to be triggered. The presence of such *ad hoc* matter fields detracts from the simplicity of the purely geometrical theory and perhaps indicates that the geometrical component is not truly fundamental. An interesting exception, however, could be the extended supergravities where certain matter-like fields are justified by an underlying supersymmetry [7] and possibly the related geometry of superspace.

Assuming that the compactification is achieved, what are the implications of taking the extra dimensions seriously? By expanding the fields in a Kaluza–Klein theory in terms of normal modes on the compact space B^K, the extra dimensions make their appearance as massive multiplets of the associated symmetry group. In this paper we wish to examine these multiplets and the higher massive excitations.

To illustrate the emergence of masses, consider the original theory of Kaluza and Klein which involves one new co-ordinate, y. The 5-dimensional line element was assumed to take form

$$ds^2 = (dx^m e_m{}^a(x))^2 - (dy - dx^m \kappa A_m(x))^2,$$

where m, n, a, b,... take the values 0, 1, 2, 3 and $e_m{}^a(x)$ represents the vierbein field on 4-dimensional spacetime. The 4-vector $A_m(x)$ is to be interpreted as the electromagnetic potential. Gauge transformations here take the form of coordinate transformations

$$x^m \to x^m,$$

$$y \to y + \Lambda(x),$$

with

$$e_m{}^a \to e_m{}^a,$$

$$A_m \to A_m + \frac{1}{\kappa}\partial_m \Lambda,$$

which leave the 1-forms, $dx^m e_m{}^a$ and $dy - dx^m \kappa A_m$, invariant. The invariant coupling to a complex scalar field, $\phi(x, y)$, would be described by the Lagrangian

$$\det e[|e_a{}^m(\partial_m \phi + \kappa A_m \partial_y \phi)|^2 - |\partial_y \phi|^2 - \mu^2 |\phi|^2].$$

If one suppose that the y-dependence of ϕ is such that

$$\partial_y \phi = -\frac{i}{R}\phi$$

then this Lagrangian reduces to that for a charged scalar,

$$\det e \left[|e_a{}^m(\partial_m \phi - ieA_m \phi)|^2 - \left(\mu^2 + \frac{1}{R^2}\right)|\phi|^2\right],$$

where the charge is given by

$$e = \frac{\kappa}{R}.$$

The electric charge is hereby understood in terms of the radius, R, of the extra dimension, interpreted as a circle. Conversely, given e, this equation fixes R (the size of the manifold) as κ/e where κ is Planck length. This also implies that the masses are in Planck units $\sim\kappa^{-1} \sim 10^{19}$ GeV. Because of this one is naturally interested to discover any exceptional cases where, for reasons of symmetry, the mass contribution is suppressed. An example is the vector A_m itself which describes a massless particle, the photon. The gauge symmetry here associated with the transformation $y \to y + \Lambda(x)$ prevents it from acquiring any mass.

To consider non-Abelian symmetries one has to interpret the new co-ordinates y^μ, $\mu = 1, 2, \ldots, \kappa$ as a parametrization of the manifold of a compact non-Abelian group, G. The line element is now assumed to take the form [4]

$$ds^2 = (dx^m e_m{}^a(x))^2 - ((dy^\mu - dx^m \kappa A_m{}^\alpha(x)\, K_\alpha{}^\mu(y))\, e_\mu{}^\beta(y))^2,$$

where the auxiliary functions $e_\mu{}^\beta(y)$ constitute an orthonormal K-bein on the group manifold G, and the functions $K_\alpha{}^\mu(y)$ are Killing vectors on G. This formalism will be reviewed in some detail in the following sections.

A further generalization is possible where the new co-ordinates are taken to parametrize some homogeneous space, G/H, on which G can act [8, 9]. A novel aspect of this kind of scheme concerns the problem of embedding. In order to give a G-invariant meaning to the ground state it is necessary to associate the motions of G/H with frame rotations (or tangent space transformations). This requires that the stability group, H, be embedded in the tangent space group, $SO(K)$, in such a way that the K-vector of $SO(K)$ has the H-content of G/H.

In all these theories the metric field carries an infinite number of new degrees of freedom corresponding to the propagation of excitations in the new dimensions. As a rule these excitations are ignored because of their large masses. However, the massive excitations are not always ignorable: their contributions to a typical scattering amplitude are comparable to those due to graviton exchange. To see this, consider the scattering of charged scalar particles with the interactions represented schematically by

$$\mathscr{L}_{\text{int}} \sim \kappa h_{mn} \partial_m \phi^* \partial_n \phi + ieA_m \phi^* \overleftrightarrow{\partial}_m \phi + \cdots,$$

where h_{mn} denotes the graviton field and A_m one of the massive excitations. The amplitude for graviton exchange is of order $\kappa^2 E^2$, while that for A-exchange is of order $e^2 E^2/M^2$, where E represents the energy scale of the scattering process. Since $M \sim e/\kappa$ these contributions are of the same order.[1] Only in the region of low momentum transfer, $t \ll E^2$, does the graviton amplitude, $\sim\kappa^2 E^4/t$, become dominant.

[1] This comparability of contributions may have relevance to the problem of renormalizability of Kaluza–Klein theories.

The purpose of this study is to deal with some of the general features of Kaluza–Klein theories. These include the geometry of homogeneous spaces, the transformations of fields defined over them, the structure of normal mode expansions[2] on G/H, the embedding of Yang–Mills gauge transformations in the group of general co-ordinate transformations, and the extraction of an effective 4-dimensional Lagrangian. Such an effective action will be particularly useful for low energy effects ascribable to the existence of the microscopic and as yet unresolvable structure on space-time. Since much of the recent work in this field has been couched in the terminology of fibre bundles [10] it is not accessible to those without expertise in such matters. We shall therefore adopt a less sophisticated approach and express results in the more generally familiar component notation.

The plan of the paper is as follows. Some definitions, notational conventions and a brief description of the $4+K$-dimensional Einstein theory are given in Section II. There follows in Section III a discussion of symmetry expects of the compactification of the extra K dimensions. Harmonic expansions on G/H are introduced to describe the excitations. An important ingredient here is the embedding of the stability group, H, in the tangent space group, $SO(K)$. Integration over G/H is used to reduce the Lagrangian to effective 4-dimensional form. Some of this is illustrated in Section IV where the low energy sector of the $4+K$-dimensional theory is shown to contain the 4-dimensional Einstein and Yang–Mills terms. Much of the pedagogical detail is confined to the Appendices, except for Appendix V which deals with the spectrum of $4+1$ Kaluza–Klein theory.

II. Extended Gravity

Let the $4+K$-dimensional spacetime be parametrized, at least in some local region, by a set of co-ordinates, z^M. On this space let there be defined at each point a local frame of reference in the form of $4+K$ linearly independent covariant vectors, $E_M{}^A$.

The basic symmetries of the theory are twofold. Firstly, there must be general covariance under the coordinate transformations, $z^M \to z'^M$. Being covariant vectors, the $E_M{}^A$ transform according to

$$E_M{}^A(z) \to E'^A_M(z') = \frac{\partial z^N}{\partial z'^M} E_N{}^A(z). \tag{2.1}$$

Secondly, there is to be covariance under a group of z-dependent linear transformations among the vectors of the reference frame,

$$E_M{}^A(z) \to E'^A_M(z) = E_M{}^B(z)\, a_B{}^A(z). \tag{2.2}$$

[2] Throughout this paper B^K is assumed to be a quotient space. One may eventually contemplate internal manifolds (associated with symmetry breaking) which are less symmetrical.

In order to establish a pseudo-Riemannian geometry on the $4 + K$-space it is usual to assume that these transformations belong to the pseudo-orthogonal group, $SO(1, 3 + K)$. We shall adhere to this choice.

With frame rotations belonging to $SO(1, 3 + K)$ one is led to consider the frame-independent combinations,

$$g_{MN} = E_M{}^A E_N{}^B \eta_{AB}, \tag{2.3}$$

which define the covariant components of the metric tensor. By an appropriate choice of basis, the $SO(1, 3 + K)$ invariant tensor, η_{AB}, can be represented by the diagonal matrix,

$$\eta_{AB} = \text{diag}(1, -1, -1, ..., -1). \tag{2.4}$$

Indices are raised and lowered in the usual way. The contravariant quantities $E_A{}^M$ and g^{MN} are defined as matrix inverses of $E_M{}^A$ and g_{MN}, respectively. (We employ the alphabetic convention that letters $M, N, ...$ from the middle alphabet will be world indices while those from the early alphabet, $A, B, ...$ shall be frame labels.)

The coupling of extended gravity to matter is conveniently illustrated by the case of the fundamental spinor of $SO(1, 3 + K)$. This spinor, $\psi(z)$, has $2^{2+[K/2]}$ components where $[K/2]$ denotes the largest integer $\leqslant K/2$. Under frame rotations the components of ψ transform as an $SO(1, 3 + K)$ spinor,

$$\psi(z) \to \psi'(z) = S(a^{-1})\, \psi(z), \tag{2.5}$$

but, under general co-ordinate transformations they are scalars,

$$\psi(z) \to \psi'(z') = \psi(z). \tag{2.6}$$

The partial derivatives $\partial_M \psi$ make a covariant world vector but they do not transform homogeneously under frame rotations. The latter defect is cured by introducing the spin connections $B_M(z)$, which constitute a gauge field with respect to frame rotations,

$$B_M \to S B_M S^{-1} + S \partial_M S^{-1} \tag{2.7}$$

and a covariant vector with respect to co-ordinate transformations. The covariant derivative is then defined by

$$\nabla_M \psi = (\partial_M + B_M)\, \psi. \tag{2.8}$$

The spin connections belong to the infinitesimal algebra of $SO(1, 3 + K)$. In the spinor representation one can write

$$B_M = \tfrac{1}{8} B_{M[AB]} [\gamma^A, \gamma^B], \tag{2.9}$$

where γ^A is a generalized Dirac matrix, i.e.,

$$\{\gamma^A, \gamma^B\} = 2\eta^{AB}. \tag{2.10}$$

With these fields it is possible to construct a suitable Lagrangian density, one which is invariant under frame rotations and a scalar density under co-ordinate transformations

$$\mathscr{L}_\psi = \frac{i}{2} \det(E)\, \bar{\psi}\gamma^A E_A{}^M \nabla_M \psi + \text{h.c.} \tag{2.11}$$

This much is nothing but a straightforward generalization to $4 + K$ dimensions of the well-known 4-dimensional structures. However, when one comes to look for the Yang–Mills potentials among the components of E and B, one will find some less familiar things. For example, the "minimal coupling" of the vectors is contained partly in the term $\bar{\psi}\gamma^A E_A{}^M \partial_M \psi$ and partly in $\bar{\psi}\gamma^A E_A{}^M B_M \psi$. The latter term may also contain a non-minimal, Pauli type of coupling.

Out of E and B and their first derivatives it is possible to construct two distinct covariant objects, or "field strengths." These are the torsion,

$$T_{MN}{}^A = \partial_M E_N{}^A - \partial_N E_M{}^A + E_N{}^C B_{MC}{}^A - E_M{}^C B_{NC}{}^A, \tag{2.12}$$

and the curvature,

$$R_{MN[AB]} = \partial_M B_{N[AB]} - \partial_N B_{M[AB]} - B_{M[AC]} B_{N[CB]} + B_{N[AC]} B_M{}[CB] \tag{2.13}$$

(where early alphabet labels are raised, lowered and contracted with η_{AB} defined in (2.4)).

The curvature scalar (density) is given by

$$\det(E)\, R = \det(E)\, E_A{}^M E_B{}^N R_{MN[AB]} \tag{2.14}$$

and can serve as a Lagrangian density for the extended gravity fields. The fields E and B may be varied independently or, alternatively, one may impose the torsion constraints, $T_{MN}{}^A = 0$, and eliminate B as an independent variable. The constraints $T = 0$ are algebraic in B and can easily be solved. One finds

$$B_{M[BC]} = \tfrac{1}{2} E_M{}^A (\Omega_{[AB]C} - \Omega_{[BC]A} + \Omega_{[CA]B}), \tag{2.15}$$

where

$$\Omega_{[AB]C} = E_A{}^M E_B{}^N (\partial_M E_{NC} - \partial_N E_{MC}). \tag{2.16}$$

III. Compactification

Having set up an action principle in $4+K$ dimensions one might inquire into the nature of the solutions to the resulting equations of motion. In a purely geometrical theory one would expect the ground state, or vacuum, to exhibit maximal symmetry. Indeed, one should find the $4+K$-dimensional Minkowski space. However, in the presence of matter fields—scalars and vectors— the situation is more complicated and it is possible to imagine that the ground state will manifest some lower symmetry. In particular, if the theory is supersymmetric the fields which accompany E and B may trigger a compactification of some of the dimensions in the ground state. We shall assume that this happens and that the ground state geometry factorizes into the product of 4-dimensional Minkowski space with a compact homogeneous space, G/H, of K dmensions. This space is invariant under the action of some group G which will be interpreted as an "internal" symmetry. We shall further assume that the compact space is small enough (size $\sim\kappa$) to justify, for most purposes, the neglect of any excitations associated with it (i.e., that such excitations would be too massive to be accessible at present energies). As mentioned in the Introduction the excitations may not always be negligible. Their contributions are generally comparable to those of gravitons and they would certainly be important if the gravitons are quantized. The aim at this stage, however, is to develop an effective 4-dimensional theory for the long range degrees of freedom.

One way to obtain the 4-dimensional theory is to expand all field variables in complete sets of harmonics on the homogeneous space G/H. The expansion coefficients are 4-dimensional fields belonging to a sequence of representations of G. Their dynamics is governed by a 4-dimensional action functional obtained by integrating out the G/H-dependence and using the orthogonality properties of harmonics on G/H. The resulting masses and couplings are necessarily G-invariant. The scale of the masses is determined by the size of the compact manifold, and we shall assume that all non-vanishing masses are therefore large. One would like to be able to extract the zero modes and discard the rest. We shall consider some aspects of this problem in the following but do not obtain a general solution. We now develop the necessary formalism.

To distinguish the compact dimensions from the remaining four of ordinary spacetime it is convenient to adopt some notational conventions. For the co-ordinates write

$$z^M = (x^m, y^\mu) \tag{3.1}$$

with middle alphabet Latin suffices, $m, n,...$ taking the values 0, 1, 2, 3 and their Greek counterparts, $\mu, \nu,...$ taking the values 1, 2,... K. The compact manifold, G/H, is parametrized by y^μ. The early alphabet frame labels are likewise divided into lower case Latin and Greek.

To implement the parametrization of G/H there are available standard group-theoretical techniques [11], one version of which we outline briefly. Suppose the co-ordinates, y^μ, label the cosets of G with respect to the subgroup, H. That is, from

each coset let there be chosen a representative element, L_y. Multiplication from the left by an arbitrary element $g \in G$ will generally carry L_y into another coset, one for which the representative element is $L_{y'}$. This defines the so-called left translation

$$gL_y = L_{y'}h, \tag{3.2}$$

where h is an element of H. Both y'^{μ} and h are determined by this equation as function of y^{μ} and g. The explicit form of these functions depends of course, on the choice of elements L_y which represent the cosets. Such details are not usually important and, here, we shall be able to manage with some general properties of the parametrization. Among these, the most important is the notion of a covariant basis on G/H.

To define a covariant basis consider the 1-form, or differential,

$$e(y) = L_y^{-1} dL_y. \tag{3.3}$$

This object belongs to the infinitesimal algebra of G and therefore can be expressed as a linear combination of the generators, $Q_{\hat{\alpha}}$,

$$\begin{aligned} e(y) &= e^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}} \\ &= dy^{\mu} e_{\mu}{}^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}}. \end{aligned} \tag{3.4}$$

The generators $Q_{\hat{\alpha}}$ fall in two categories: the set $Q_{\bar{\alpha}}$ which generates the subgroup, H, and the remainder, Q_{α}, $\alpha = 1, 2, \ldots, K$, associated with the cosets, G/H. Correspondingly, one writes

$$e(y) = e^{\alpha}(y)\, Q_{\alpha} + e^{\bar{\alpha}}(y)\, Q_{\bar{\alpha}}. \tag{3.5}$$

Of special interest are the coefficients $e_{\mu}{}^{\alpha}(y)$ which constitute a nonsingular $K \times K$ matrix. The columns of this matrix provide the covariant components of the K linearly independent vectors of the covariant basis. The contravariant components of the reciprocal basis, $e_{\alpha}{}^{\mu}$, are defined by inverting the matrix $e_{\mu}{}^{\alpha}$.

The behaviour of $e(y)$ under left translations can be deduced from (3.2) and (3.3). Thus, quite generally,

$$\begin{aligned} e(y) \to e(y') &= hL_y^{-1} g^{-1} d(gL_y h^{-1}) \\ &= he(y)\, h^{-1} + hdh^{-1} + hL_y^{-1} g^{-1} dg L_y h^{-1}. \end{aligned} \tag{3.6}$$

On substituting the expansion in terms of generators and introducing the matrices, $D_{\hat{\alpha}}{}^{\hat{\beta}}$, of the adjoint representation of G, defined by

$$g^{-1} Q_{\hat{\alpha}}\, g = D_{\hat{\alpha}}{}^{\hat{\beta}}(g)\, Q_{\hat{\beta}} \tag{3.7}$$

the formula (3.6) gives

$$e^{\hat{\alpha}}(y') = e^{\hat{\beta}}(y)\, D_{\hat{\beta}}{}^{\hat{\alpha}}(h^{-1}) + (hdh^{-1})^{\hat{\alpha}} + (g^{-1}dg)^{\hat{\beta}}\, D_{\hat{\beta}}{}^{\hat{\alpha}}(L_y h^{-1}). \tag{3.8}$$

The components, $(hdh^{-1})^{\hat{\alpha}}$ and $(g^{-1}dg)^{\hat{\alpha}}$, of elements in the algebra are naturally defined by

$$g^{-1}dg = (g^{-1}dg)^{\hat{\alpha}}\, Q_{\hat{\alpha}},$$
$$hdh^{-1} = (hdh^{-1})^{\hat{\alpha}}\, Q_{\hat{\alpha}} = (hdh^{-1})^{\bar{\alpha}}\, Q_{\bar{\alpha}}. \tag{3.9}$$

Since hdh^{-1} belongs to the algebra of H, we must have $(hdh^{-1})^{\alpha} = 0$.

(A more detailed discussion of the properties of the "boost" elements, L_y, together with some illustrations is given in Appendix I. Curvature and torsion formulae for G/H are derived in Appendix II, and various implications of the transformation formula (3.8) are derived in Appendix III.)

Now consider the question of harmonic expansions on the internal space. To begin with, if the internal space were simply the group G itself (i.e., $H = 1$) it would be natural to employ the full set of matrices of the unitary irreducible representations of G. For the function, $\phi(g)$, one would write

$$\phi(g) = \sum_n \sum_{p,q} \sqrt{d^n}\, D^n_{pq}(g)\, \phi^n_{qp}. \tag{3.10}$$

where D^n_{pq} is a unitary matrix of dimension d_n and the sum includes all matrix elements of all the unitary irreducible representations, $g \to D^n(g)$. The coefficients, ϕ^m_{pq}, are projected out by integrating over the group,

$$\phi^n_{pq} = \frac{\sqrt{d_n}}{V_G} \int_G d\mu D^n_{pq}(g^{-1})\, \phi(g), \tag{3.11}$$

where $d\mu$ is the invariant measure normalized to volume V_G.

For functions on the coset space G/H the expansion is somewhat restricted. Typically, one is concerned with functions $\phi_i(g)$ that are subject to the auxiliary symmetry

$$\phi_i(hg) = \mathbb{D}_{ij}(h)\, \phi_j(g), \tag{3.12}$$

where $h \in H$ and $\mathbb{D}(h)$ is some particular representation of H. This means that $\phi(g)$, though not strictly constant over the points in a coset, is related by a linear rule, i.e.,

$$\phi_i(g_1) = \mathbb{D}_{ij}(g_1 g_2^{-1})\, \phi_j(g_2)$$

for g_1 and g_2 in the same coset. The appropriate restriction of the expansion (3.10) is clear. It must include only those terms for which

$$\mathbb{D}^n(hg) = \mathbb{D}(h)\, D^n(g).$$

For irreducible $\mathbb{D}(h)$ one should write

$$\phi_i(g) = \sum_n \sum_{\zeta,q} \sqrt{\frac{dn}{d_{\mathbb{D}}}}\, D^n_{i\zeta,q}(g)\, \phi^n_{q\zeta}, \tag{3.13}$$

where the sum includes all unitary irreducible representations of G for which $D^n_{i\zeta,q}$ makes sense, i.e., all those which include $\mathbb{D}(h)$ on restriction to H. If $\mathbb{D}(h)$ is contained more than once in a given $D^n(g)$ then a supplementary label, ζ, is needed to distinguish them. The dimension of $\mathbb{D}(h)$ is denoted $d_{\mathbb{D}}$. The coefficients, $\phi^n_{q\zeta}$, are projected from the sum (3.13) by integrating over the coset space,

$$\phi^n_{q\zeta} = \frac{1}{V_K}\sqrt{\frac{d_n}{d_{\mathbb{D}}}}\int_{G/H} d\mu D^n_{q,i\zeta}(L_y)\,\phi_i(L_y^{-1}). \tag{3.14}$$

This expression can be derived from (3.11) by resolving the general element g^{-1} into the product of appropriate elements L_y and h. Such a decomposition is unique once the form of the representative element L_y has been chosen. From (3.12) it follows that the h-dependence factorizes,

$$\begin{aligned}\phi_i(g) &= \phi_i(h^{-1}L_y{}^{-1}) \\ &= \mathbb{D}_{ij}(h^{-1})\,\phi_j(L_y{}^{-1}).\end{aligned}$$

Integration over the subgroup H can be carried out explicitly and leaves the result (3.14).

Thus, for a set of fields $\psi_i(x,y)$ the general form of the harmonic expansion on G/H is given by

$$\psi_i(x,y) = \sum_n \sum_{\zeta,q} \sqrt{\frac{d_n}{d_{\mathbb{D}}}}\, D^n_{i\zeta,q}(L_y{}^{-1})\,\psi^n_{q\zeta}(x). \tag{3.15}$$

Here n labels the exitations.

The coefficients fields $\psi^n_{q\zeta}(x)$ are obtained by integrating over G/H as in (3.14). If the representation $\mathbb{D}(h)$ to which ψ_i belongs is reducible then it would be necessary to separate ψ_i into irreducible parts and make an expansion of the form (3.15) for each. This raises the important question of how the H content of $\psi(x,y)$ is to be determined. It will be settled by invoking the ground state symmetry.

The full set of fields in $4+K$ dimensions is supposed to be governed by an action functional which is invariant with respect to both the general co-ordinate transformations and the frame rotations. These symmetries will not be manifest in the ground state, however. By assumption the ground state geometry is $M^4 \times G/H$, where G/H is a compact K-dimensional manifold. In the ground state we therefore have

$$\langle E_M{}^A(x,y)\rangle = \begin{pmatrix} \delta^a_m & 0 \\ 0 & e_\mu{}^\alpha(y) \end{pmatrix}, \tag{3.16}$$

where $e_\mu{}^\alpha$ is the covariant basis on G/H defined above. The expression (3.16) is invariant under the product of the 4-dimensional Poincaré group and the local (x-dependent) transformations of G. The Poincaré invariance is exactly as in the familiar 4-dimensional theory of general relativity. The local G-invariance is compounded of the x-dependent left translations on G/H, $y \to y'(x,y)$ and the induced tangent space

transformations, $D_\alpha{}^\beta(h)$, with h defined by (3.2). To see this, consider the action of the combined transformation on $E_\mu{}^\alpha(x, y)$

$$E_\mu{}^\alpha(x,y) \to E'^\alpha_\mu(x,y') = \frac{\partial y^\nu}{\partial y'^\mu} E_\nu{}^\beta(x,y)\, D_\beta{}^\alpha,$$

where $D_\beta{}^\alpha$ is a K-dimensional orthogonal matrix. In order to have

$$\langle E'^\alpha_\mu(x,y)\rangle = \langle E_\mu{}^\alpha(x,y)\rangle$$

it is only necessary to choose $\langle E_\mu{}^\alpha\rangle = e_\mu{}^\alpha$ and $D_\beta{}^\alpha = D_\beta{}^\alpha(h)$. This follows from (3.8) which contains the formula

$$dy'^\mu e_\mu{}^\alpha(y') = dy^\mu e_\mu{}^\beta(y)\, D_\beta{}^\alpha(h^{-1}).$$

Thus the embedding is fixed. With the left translation, $y \to y'$, must be associated the $SO(K)$ transformation, $D_\alpha{}^\beta(h)$, where $h \in H$. This implies that the K-vector of $SO(K)$ has the H-content of G/H. More precisely, the generators, $Q_{\bar\alpha}$, of H can be expressed in terms of the generators, $\Sigma^{\alpha\beta}$, of $SO(K)$ by the formula[3]

$$Q_{\bar\gamma} = -\tfrac{1}{2} c_{\bar\gamma\alpha\beta}\Sigma^{\alpha\beta}, \tag{3.17}$$

where the coefficients $c_{\bar\gamma\alpha\beta}$ are taken from the set of structure constants $c_{\gamma\alpha\beta}$ of G.

The Yang–Mills group of transformations described here, which leaves the ground state invariant also preserves the form of the expansion (3.15). Thus, with ψ transforming according to

$$\psi_i(x,y) \to \psi'_i(x,y') = \mathbb{D}_{ij}(h)\,\psi_j(x,y) \tag{3.18}$$

it is a simple matter to derive the corresponding rule for the expansion coefficients

$$\psi^n_{p\zeta}(x) \to \psi'^n_{p\zeta}(x) = D^n_{pq}(g)\,\psi^n_{q\zeta}(x). \tag{3.19}$$

They belong to the various irreducible representations of G which contribute to the expansion (3.15).

In order to express the full $4+K$-dimensional theory in its equivalent 4-dimensional form, one should substitute expansions like (3.15) for all of the fields in the system. The 4-dimensional Lagrangian is then obtained by integrating the $4+K$ dimensional one over the compact manifold, G/H. Integration of products of matrices

[3] For an infinitesimal transformation in H we have

$$D_{\alpha\beta}(h) = \delta_{\alpha\beta} + \delta h^{\bar\gamma} c_{\alpha\bar\gamma\beta}$$
$$= \delta_{\alpha\beta} + \omega_{\alpha\beta},$$

where $\omega_{\alpha\beta} = -\omega_{\beta\alpha}$. Viewed as an infinitesimal $SO(K)$ transformation this reads $(1/2)\,\omega_{\alpha\beta}\Sigma^{\alpha\beta} = \delta h^{\bar\gamma} Q_{\bar\gamma}$, which implies (3.17).

$D^n(L_y^{-1})$ reduces, in principle, to an algebraic problem. Although it is difficult to obtain results without going into the details of particular cases, the most important feature, orthogonality, can be illustrated by an example.

Consider the invariant integral

$$I^{nn'}_{p\zeta,p'\zeta'} = \int_{G/H} d^K y \det e(y)\, D^n_{p,i\zeta}(L_y)\, D^{n'}_{i\zeta',p'}(L_y{}^{-1}), \tag{3.20}$$

where the label, i, is understood to be summed. Since the measure, $d\mu = d^K y$ det e, is invariant under left translations, one can write

$$\begin{aligned} I^{nn'}_{p\zeta,p'\zeta'} &= \int d\mu D^n_{p,i\zeta}(L_{y'})\, D^{n'}_{i\zeta',p'}(L_{y'}{}^{-1}) \\ &= \int d\mu D^n_{p,i\zeta}(gL_y h^{-1})\, D^{n'}_{i\zeta',p'}(hL_y{}^{-1}g^{-1}) \\ &= D^n_{pq}(g)\, I^{nn'}_{q\zeta,q'\zeta'} D^{n'}_{q'p'}(g^{-1}) \end{aligned}$$

because the dependence on h cancels. Hence I is an invariant tensor of G. By Schur's lemma it is must take the form

$$I^{nn'}_{p\zeta,p'\zeta'} = \delta_{nn'}\delta_{pp'}k^n_{\zeta\zeta'}.$$

To evaluate $k^n_{\zeta\zeta'}$, set $n = n'$, $p = p'$ and sum over p. One obtains

$$\begin{aligned} d_n k^n_{\zeta\zeta'} &= \int d\mu D^n_{i\zeta',p}(L_y^{-1})\, D^n_{p,i\zeta}(L_y) \\ &= V_K d_{\mathrm{D}} \delta_{\zeta\zeta'}, \end{aligned}$$

where d_n and d_{D} are the dimensions of D^n_{pq} and $\mathbb{D}_{ij}$, respectively. Thus

$$I^{nn'}_{p\zeta,p'\zeta'} = V_K \frac{d_{\mathrm{D}}}{d_n} \delta_{\zeta\zeta'}\delta_{pp'}\delta_{nn'}. \tag{3.21}$$

This result is needed for inverting the expansion (3.15).

The argument leading to (3.21) generalizes. For example, one can write

$$\begin{aligned} &\int_{G/H} d\mu c^{i_1 i_2 i_3} D^{n_1}_{i_1\zeta_1,p_1}(L_y^{-1})\, D^{n_2}_{i_2\zeta_2,p_2}(L_y^{-1})\, D^{n_3}_{i_3\zeta_3,p_3}(L_y^{-1}) \\ &\quad = a_{n_1 n_2 n_3} V_K c^{i_1 i_2 i_3} \begin{pmatrix} n_1 & n_2 & n_3 \\ i_1\zeta_1 & i_2\zeta_2 & i_3\zeta_3 \end{pmatrix} \begin{pmatrix} n_1 & n_2 & n_3 \\ p_1 & p_2 & p_3 \end{pmatrix}, \end{aligned}$$

where $c^{i_1 i_2 i_3}$ is an invariant tensor of H and the 3–n symbols are essentially Clebsch–Gordan coefficients for the unitary representations of G. The factor $a_{n_1 n_2 n_3}$ depends on normalization conventions for the 3–n symbols.

Having given, in outline,[4] the general form of the expansions to be used, we shall in the following section restrict our considerations to the leading terms. The justification for this restriction is simply that, in the regime of low energies, only zero-modes are important.

IV. Effective Lagrangians

It is guaranteed that the graviton and a set of Yang–Mills vectors are included among the zero-mass fields of the 4-dimensional effective theory. This results from the gauge symmetries of the $4+K$-dimensional theory and the assumed symmetry of the ground state. It can be proved as follows. For $E_M{}^A(x,y)$ consider the ansatz [9]

$$E_M{}^A(x,y)=\begin{pmatrix} E_m{}^a(x) & -A_m{}^\beta(x)\,D_\beta{}^\alpha(L_y) \\ 0 & e_\mu{}^a(y) \end{pmatrix}, \tag{4.1}$$

where the y-dependence is specified. This ansatz is compatible with a subgroup of the $4+K$-dimensional symmetries: the 4-dimensional general co-ordinate transformations,

$$x^m \to x'^m(x)$$

and the Yang–Mills transformations, or x-dependent left translations

$$y^\mu \to y'^\mu(x,y)$$

with associated frame rotations, $D_\alpha{}^\beta(h)$,

$$E'^a_m(x')=\frac{\partial x^n}{\partial x'^m}E_n{}^a(x), \tag{4.2a}$$

$$E'^\alpha_m(x',y')=\left(\frac{\partial x^n}{\partial x'^m}E_n{}^\beta(x,y)+\frac{\partial y^\mu}{\partial x'^m}e_\mu{}^\beta(y)\right)D_\beta{}^\alpha(h^{-1}), \tag{4.2b}$$

$$e_\mu{}^\alpha(y')=\frac{\partial y^\nu}{\partial y'^\mu}e_\nu{}^\beta(y)\,D_\beta{}^\alpha(h^{-1}). \tag{4.2c}$$

Of these rules, (4.2a) requires no comment and (4.2c) has been dealt with already in Section III. There remains (4.2b) which implies, in view of the ansatz,

$$-A'^\beta_m(x')\,D_\beta{}^\alpha(L_{y'})=\left(-\frac{\partial x^n}{\partial x'^m}A_n{}^{\hat\gamma}(x)\,D_{\hat\gamma}{}^\beta(L_y)+\frac{\partial y^\mu}{\partial x'^m}e_\mu{}^\beta(y)\right)D_\beta{}^\alpha(h^{-1}).$$

[4] Some illustrative examples are contained in Appendix IV.

With the help of formulae derived in Appx. 3 this rule reduces to the form

$$A'_m(x') = \frac{\partial x^n}{\partial x'^m}\,(gA_n(x)\,g^{-1} - g\partial_n\,g^{-1}), \tag{4.3}$$

where $A_m = A_m{}^{\hat{\alpha}} Q_{\hat{\alpha}}$. This is precisely the rule to be expected for a Yang–Mills potential.

The zero torsion spin connections corresponding to the ansatz (4.1) are obtained by substituting in the formulae (2.15), (2.16)

$$\begin{aligned}
&B_{a[bc]} = \tfrac{1}{2}E_{[a}{}^m E_{b]}{}^n\,\partial_m E_{nc} - \tfrac{1}{2}E_{[b}{}^m E_{c]}{}^n\,\partial_m E_{na} + \tfrac{1}{2}E_{[c}{}^m E_{a]}{}^n\,\partial_m E_{nb},\\
&B_{a[b\gamma]} = -B_{\gamma[ab]} = \tfrac{1}{2}E_a{}^m E_b{}^n F_{mn}{}^{\hat{\alpha}} D_{\hat{\alpha}\gamma}(L_y),\\
&B_{a[\beta\gamma]} = A_a{}^{\bar{\delta}}(D_{\bar{\delta}}{}^{\bar{\alpha}}(L_y) - D_{\bar{\delta}}{}^{\alpha}(L_y)\,\pi_\alpha{}^{\bar{\alpha}}(y))\,c_{\bar{\alpha}\beta\gamma},\\
&B_{\alpha[\beta c]} = 0,\\
&B_{\alpha[\beta\gamma]} = \tfrac{1}{2}c_{\alpha\beta\gamma} + \pi_\alpha{}^{\bar{\alpha}}(y)\,c_{\bar{\alpha}\beta\gamma}.
\end{aligned} \tag{4.4}$$

In these expressions the following notations are used:

$$\begin{aligned}
E_{[a}{}^m E_{b]}{}^n &= E_a{}^m E_b{}^n - E_b{}^m E_a{}^n,\\
F_{mn}{}^{\hat{\alpha}} &= \partial_m A_n{}^{\hat{\alpha}} - \partial_n A_m{}^{\hat{\alpha}} - A_m{}^{\hat{\beta}} A_n{}^{\hat{\gamma}} c_{\hat{\gamma}\hat{\beta}}{}^{\hat{\alpha}},\\
\pi_\alpha{}^{\bar{\alpha}}(y) &= e_\alpha{}^\mu(y)\,e_\mu{}^{\bar{\alpha}}(y),
\end{aligned} \tag{4.5}$$

and the structure constants of G are defined by

$$[Q_{\hat{\alpha}}, Q_{\hat{\beta}}] = c_{\hat{\alpha}\hat{\beta}}{}^{\hat{\gamma}} Q_{\hat{\gamma}}.$$

On substituting the expressions (4.4) into the $4 + K$-dimensional curvature scalar one obtains

$$R_{4+K} = R_4 - \tfrac{1}{4}F_{ab}{}^{\hat{\alpha}} F_{ab}{}^{\hat{\beta}} D_{\hat{\alpha}}{}^{\gamma}(L_y)\,D_{\hat{\beta}}{}^{\gamma}(L_y) + R_K, \tag{4.6}$$

where $R_4 = R_4(E)$ is the usual 4-dimensional curvature scalar, expressed in terms of the vierbein $E_m{}^a(x)$, and R_K is the constant curvature of G/H (see Appendix II). On integrating y^μ the Yang–Mills Lagrangian emerges, since

$$\frac{1}{V_K}\int d^K y\,\det e(y)\,D_{\hat{\alpha}}{}^{\gamma}(L_y)\,D_{\hat{\beta}}{}^{\gamma}(L_y) = k\delta_{\hat{\alpha}\hat{\beta}},$$

where $k = \dim(G/H)/\dim(G)$. Thus, the above assertion that the system contains the graviton and massless Yang–Mills fields is verified.

There may also be zero mass scalars in the system, like the Brans–Dicke scalar in the $K = 1$ theory. This point is sensitive, however, to the matter field couplings which

are generally needed to force the compactification. We shall not therefore discuss the scalar modes.

Finally, to illustrate how the ansatz described above can be combined with an analogous one for a spinor field to give an effective 4-dimensional term, we consider the expression

$$\mathscr{L}_* = \frac{i}{2} \det E(x,y)\, \bar{\psi}\gamma^A E_A{}^M \nabla_M \psi + \text{h.c.} \tag{4.7}$$

with $\psi(x,y)$ represented by the ansatz,

$$\psi(x,y) = D(L_y{}^{-1})\, \psi(x), \tag{4.8}$$

This is a typical term from the harmonic expansion of $\psi(x,y)$. The spinor of $SO(1,3+K)$ must of course be decomposed relative to the subgroup $SO(1,3) \times H$; each piece in the decomposition will yield a $\psi(x)$ and the details of labelling are ignored here. One must substitute the expressions (4.1), (4.4) and (4.8) into (4.7),

$$\begin{aligned}\mathscr{L}_* = \frac{i}{2} \det E(x) \det e(y)\, \bar{\psi}(x)\, D(L_y) \Big[\gamma^a \Big\{ E_a{}^m \partial_m + E_a{}^\mu \partial_\mu - \frac{1}{4} B_{a[BC]} \gamma^{BC} \Big\} \\ + \gamma^\alpha \Big\{ e_\alpha{}^\mu \partial_\mu - \frac{1}{4} B_{\alpha[BC]} \gamma^{BC} \Big\} \Big] D(L_y{}^{-1})\, \psi(x) + \text{h.c.}\end{aligned} \tag{4.9}$$

To simplify this expression it is necessary to eliminate the derivative ∂_μ by means of the formula

$$\begin{aligned}\partial_\mu D(L_y{}^{-1}) &= -D(L_y{}^{-1})\, \partial_\mu D(L_y)\, D(L_y{}^{-1}) \\ &= -e_\mu{}^{\dot{\alpha}}(y)\, D(Q_{\dot{\alpha}})\, D(L_y{}^{-1})\end{aligned}$$

which follows from the definition (3.3). Multiplication by $e_\alpha{}^\mu$ then gives

$$e_\alpha{}^\mu \partial_\mu D(L_y{}^{-1}) = -(D(Q_\alpha) + \pi_\alpha{}^{\bar{\alpha}} D(Q_{\bar{\alpha}}))\, D(L_y{}^{-1}),$$

where π is defined in (4.5). It turns out that the π-containing terms are cancelled by similar terms in the spin connections. This cancellation is ensured by the embedding of H in the tangent space group (see Appendix IV). After some tedious calculations it is found that (4.9) reduces to

$$\begin{aligned}\mathscr{L}_* = \frac{i}{2} \det E(x) \det e(y)\, \bar{\psi}(x)\, D(L_y) \Big[\gamma^a \Big\{ E_a{}^m(x)\, \partial_m + \frac{1}{4} B_{a[bc]}(x)\, \gamma^{bc} \\ - A_a{}^{\dot{\alpha}}(x)\, D(Q_{\dot{\alpha}}) \Big\} + \frac{3}{8} F_{ab}{}^\delta(x)\, \gamma^{ab}\gamma^a D_\alpha{}^\delta(L_y{}^{-1}) \\ - \gamma^a D(Q_\alpha) + \frac{1}{8} c_{\alpha\beta\gamma}\gamma^{\alpha\beta\gamma} \Big] D(L_y{}^{-1})\, \psi(x) + \text{h.c.}\end{aligned} \tag{4.10}$$

and it remains only to integrate over y. Since γ^a must commute with $D(L_y)$, the result takes the form

$$\frac{i}{2} V_K \det E(x)\, \bar{\psi}(x) \left[\gamma^a \left\{E_a{}^m(x)\, \partial_m + \frac{1}{4} B_{a[bc]}(x)\, \gamma^{bc} - A_a{}^{\beta}(x)\, D(Q_\beta)\right\} + \frac{3}{8} F_{ab}{}^{\delta}(x)\, \gamma^{ab} \Gamma^{\delta} - M\right] \psi(x) + \text{h.c.}, \tag{4.11}$$

where Γ^{δ} and M are matrices defined by

$$\Gamma^{\delta} = \frac{1}{V_K} \int_{G/H} d\mu D(L_y)\, \gamma^a D(L_y{}^{-1})\, D_a{}^{\delta}(L_y{}^{-1}), \tag{4.12}$$

$$M = \frac{1}{V_K} \int_{G/H} d\mu D(L_y)(\gamma^a D(Q_a) - \frac{1}{8} c_{\alpha\beta\gamma} \gamma^{\alpha\beta\gamma})\, D(L_y{}^{-1}). \tag{4.13}$$

Note that the "Pauli-moment term" implied by non-zero Γ^{δ} is a necessary consequence of our formalism, and the precise numerical expressions for Γ^{δ} and M are the hall-marks of the expansion on G/H.

V. Conclusions

In this study we have attempted to analyze in some detail various formal aspects of Kaluza–Klein theories in which the ground state geometry takes the form $M^4 \times B^K$, where B^K is a K-dimensional quotient space G/H. The symmetries of this ground state are used to classify the excitations of the system, and these excitations are found to include particles of spins zero, one and two, in general. In particular, there are a set of four vectors belonging to the adjoint representation of G contained among the components of the metric tensor and spin connection which serve as Yang–Mills potentials.

Harmonic expansions on B^K are used to define sequences of component fields, defined on M^4 and belonging to irreducible representations of the Yang–Mills group, G. An important feature in determining the content of these expansions is the embedding of the stability group H in the tangent space group $SO(K)$. This embedding is discussed in detail. Considerable use is made of the "boosts," L_y, in setting up the formalism. The transformation properties of L_y are used to define the action of G on the manifold G/H, and the derivatives of L_y give the covariant basis vectors on this manifold. They also are used in the construction of integral formulae for mass matrices and coupling parameters for Pauli-moment-like terms in the 4-dimensional effective theory (formulae (4.12) and (4.13)).

We remark that the ansatz for the 4-dimensional fields associated with massless particles—graviton, Yang–Mills vectors and Brans–Dicke-type scalars—can be seen to arise as leading terms in the harmonic expansions. The complete spectrum for the

$4+1$-dimensional theory is discussed in Appendix V. It is shown there that the massive excitations are purely of spin-2 and that they can be assigned to infinite-dimensional representations of the non-compact group $SO(1,2)$. Such non-compact symmetries are spontaneously broken and may be viewed as spectrum generating symmetries. It is interesting that, viewed in this light, Kaluza–Klein theory seems to constitute an example of a consistent theory of massive, charged spin-2 particles.

Information drawn from the structure of the harmonic expansions would be useful in testing the classical stability of Kaluza–Klein type solutions. We have not dealt with this type of investigation (except for the simple example of $4+1$ theory) since it is generally necessary to introduce matter fields of some sort into the system to generate non-vanishing curvature in G/H.[5] Such matter fields, which must affect the stability considerations, represent a non-geometrical element in the theory (except, perhaps, in the case of extended supergravitities [7]) and are particular to specific models.

APPENDIX I: Non-linear Realizations

To supplement the discussion of Section III we give here a somewhat more amplified description of the properties of non-linear realizations of continuous groups. The concepts were thoroughly exposed in the physics literature a few years ago when their appropriateness in treating spontaneously broken symmetries was noticed [12]. In that work the space on which the group acts is a set of scalar fields and the non-linearity can be expressed through algebraic constraints on these fields. Because of such constraints it is impossible for all components to vanish simultaneously, hence the absence of an invariant ground state, and hence the relevance to spontaneous symmetry breaking. From a more abstract point of view, the set of scalar fields constitute a vector space which supports a linear representation of the symmetry. The constraints serve to pick out a subspace, generally curved, which is carried into itself by the action of the group. On such a subspace the group is said to be realized non-linearly. Indeed, to characterize the fields belonging to such a curved subspace it is generally necessary to adopt a curvilinear co-ordinate system and so to express the group actions by the transformations which they induce in these co-ordinates.

The classic example of non-linear realizations is the treatment of chiral $SU(2)\times SU(2)$. This group has the infinitesimal structure of $SO(4)$ and the scalar fields (σ, π) are assumed to transform as a 4-vector. If they are subjected to the constraint

$$\sigma^2+\pi^2=F^2, \tag{I.1}$$

where F is a fixed number, then the number of independent components is reduced to three: they belong to a 3-sphere of radius F. This 3-sphere is of course carried into

[5] The original Kaluza–Klein theory in $4+1$-dimensions represents a case where G/H is flat. Generalizations of this have been investigated by Scherk and Schwarz [13].

itself by the transformations of $SO(4)$. It is invariant. To parametrize the points of the 3-sphere one could simply use the 3-vector π and write $\sigma = \pm\sqrt{F^2 - \pi^2}$. For some purposes, however, it is more useful to define a set of boosts, L_π, which are $SO(4)$ transformations that carry a given reference point, say

$$(\sigma, \pi) = (F, 0) \tag{I.2}$$

into a general point. In terms of the 4×4 orthogonal matrices $D_{\alpha\beta}(g)$ of the vector representation one writes

$$\begin{aligned} \sigma &= D_{44}(L_\pi)\, F, \\ \pi_i &= D_{i4}(L_\pi)\, F. \end{aligned} \tag{I.3}$$

Of course there is a great deal of latitude in the choice of L_π and it will generally be necessary to use at least two co-ordinate patches if the sphere is to be covered without generating artificial singularities.

To illustrate the parametrization of the sphere we give the example of spherical polar co-ordinates,

$$\begin{aligned} \sigma &= F\cos\theta, \\ \pi_1 &= F\sin\theta\cos\varphi, \\ \pi_2 &= F\sin\theta\sin\varphi\cos\psi, \\ \pi_3 &= F\sin\theta\sin\varphi\sin\psi, \end{aligned} \tag{I.4}$$

which corresponds to the choice of L_π,

$$L_\pi = e^{-\varphi Q_{23}} e^{-\varphi Q_{12}} e^{-\theta Q_{41}}. \tag{I.5}$$

Here the operator $Q_{\alpha\beta} = -Q_{\beta\alpha}$ is the generator of the infinitesimal rotations in the $\alpha\beta$-plane

$$[Q_{\alpha\beta}, Q_{\gamma\delta}] = \delta_{\beta\gamma} Q_{\alpha\delta} - \delta_{\alpha\gamma} Q_{\beta\delta} + \delta_{\alpha\delta} Q_{\beta\gamma} - \delta_{\beta\delta} Q_{\alpha\gamma}. \tag{I.6}$$

On applying an $SO(4)$ transformation, g, to the sphere, the point whose co-ordinates are (θ, φ, ψ) will be carried into one with co-ordinates $(\theta', \varphi', \psi')$. A rotation through the angle ω in the 14-plane, for example, gives

$$\begin{pmatrix} \cos\theta' \\ \sin\theta'\cos\varphi' \end{pmatrix} = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta\cos\varphi \end{pmatrix}.$$

By solving these equations one can discover the transformed boost, $L_{\pi'}$. In general the co-ordinate transformation is given by

$$\begin{aligned} D_{\alpha 4}(L_{\pi'}) &= D_{\alpha\beta}(g)\, D_{\beta 4}(L_\pi) \\ &= D_{\alpha 4}(gL_\pi). \end{aligned}$$

From this it follows that the precise relation between L_π and $L_{\pi'}$ must take the form

$$gL_\pi = L_{\pi'}h, \tag{I.7}$$

where h is an orthogonal transformation in the 123-subspace, i.e.,

$$D_{\alpha 4}(h) = \delta_{\alpha 4}. \tag{I.8}$$

Since $L_{\pi'}$ is known once the angles $(\theta', \varphi', \psi')$ have been computed, the $SO(3)$ transformation, $h = L_{\pi'}{}^{-1}gL_\pi$, is well defined.

To compute the basis vectors on the 3-sphere consider the differential

$$\begin{aligned} e &= L_\pi{}^{-1}dL_\pi \\ &= e^{\theta Q_{41}}e^{\varphi Q_{12}}e^{\psi Q_{23}}d(e^{-\psi Q_{23}}e^{-\varphi Q_{12}}e^{-\theta Q_{41}}) \\ &= -d\theta Q_{41} - d\varphi(Q_{12}\cos\theta + Q_{42}\sin\theta) \\ &\quad - d\psi(Q_{23}\cos\varphi + (Q_{13}\cos\theta - Q_{43}\sin\theta)\sin\varphi) \\ &= -d\theta Q_{41} - d\varphi\sin\theta Q_{42} + d\psi\sin\theta\sin\varphi Q_{43} \\ &\quad - d\varphi Q_{12} - d\psi\cos\varphi Q_{23} + d\psi\cos\theta\sin\varphi Q_{31}. \end{aligned} \tag{I.9}$$

From this expression one extracts the coefficients

$$e_\mu{}^\alpha = \begin{array}{c} \\ \theta \\ \varphi \\ \psi \end{array} \begin{array}{c} \begin{array}{ccc} 41 & 42 & 43 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 \\ 0 & -\sin\theta & 0 \\ 0 & 0 & \sin\theta\sin\varphi \end{pmatrix} \end{array},$$

$$e_\mu{}^{\bar\alpha} = \begin{array}{c} \\ \theta \\ \varphi \\ \psi \end{array} \begin{array}{c} \begin{array}{ccc} 12 & 23 & 31 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -\cos\varphi & \cos\theta\sin\varphi \end{pmatrix} \end{array}, \tag{I.10}$$

where the labelling of rows and columns is indicated. The metric tensor on the 3-sphere is given by

$$\begin{aligned} ds^2 &= (d\theta e_\theta{}^\alpha + d\varphi e_\varphi{}^\alpha + d\psi e_\psi{}^\alpha)^2 \\ &= d\theta^2 + \sin^2\theta\, d\varphi^2 + \sin^2\theta\sin^2\varphi\, d\psi^2 \end{aligned} \tag{I.11}$$

as would be expected in spherical polar co-ordinates. The spin connection $B_{\alpha[\beta\gamma]}$ is constructed out of the quantities

$$\pi_{\alpha}{}^{\overline{\beta}} = e_{\alpha}{}^{\mu} e_{\mu}{}^{\overline{\beta}}$$

$$= \begin{array}{c} \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{c} \begin{array}{ccc} 12 & 23 & 31 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ \dfrac{1}{\sin\theta} & 0 & 0 \\ 0 & -\dfrac{\cot\varphi}{\sin\theta} & \cot\theta \end{pmatrix} \end{array} \tag{I.12}$$

as discussed in the text, (4.4) (and further in Appendix 2).

The case discussed above is an example of a non-linear realization in which the group $G = SO(4)$ is made to act on a 3-sphere. The points of the 3-sphere are put into correspondence with the left cosets $SO(4)/SO(3)$ through the parametrization of the boosts, L_{π}. The procedure generalizes directly to manifolds of the form G/H, where H is some subgroup of G. (Right cosets can also be used. One writes

$$g: L_y \to L_{y'} = hL_y g^{-1}.$$

On the manifold of G itself one can define the action of $G \times G$

$$(g_1, g_2): L_y \to L_{y'} = g_1 L_y g_2{}^{-1}.$$

In all such cases the mapping $y \to y'$ is determined once the form of L_y has been specified.)

Global properties of a quotient space G/H are often made transparent by embedding in a flat space of higher dimension, i.e., by viewing the non-linear realization as obtained from a linear one with constraints applied. A simple illustration of this is provided by the case of $SU(3)/SU(2) \times U(1)$ which we consider briefly.

The infinitesimal transformations of $SU(3)$ are generated by the eight operators

$$Q_a{}^b = -Q_b{}^{a\dagger}, \qquad a, b = 1, 2, 3 \qquad (Q_a{}^a = 0). \tag{I.13}$$

The boosts may be given by

$$L_y = \exp(y^{\alpha} Q_{\alpha}{}^3 - \text{h.c.}), \tag{I.14}$$

where the co-ordinates y^{α}, $\alpha = 1, 2$, are complex. The little group in this case, $H = SU(2) \times U(1)$, is generated by $Q_{\alpha}{}^{\beta}$. With this choice of L_y the procedures discussed above could be used to find the mapping $y^{\alpha} \to y'^{\alpha}$ and the induced $SU(2) \times U(1)$ transformation, h, corresponding to an $SU(3)$ transformation, g. However, we shall not pursue this.

The manifold $SU(3)/SU(2) \times U(1)$ clearly has two complex dimensions. In order

to see that it is, in fact, $\mathbb{C}P_2$ one can proceed as follows. Let the hermitian, traceless matrices, $Y_a{}^b$, be the co-ordinates of a flat space on which $SU(3)$ acts linearly

$$Y \to Y' = g\, Y g^{-1}. \tag{I.15}$$

An invariant subspace can be picked out by requiring the co-ordinates to satisfy the matrix equations

$$Y^2 = \frac{R}{3} Y + \frac{2}{9} R^2, \tag{I.16}$$

where R is a constant. Since Y is hermitian, it can be diagonalized by an $SU(3)$ transformation and one finds easily that its diagonalized form must be

$$\hat{Y} = \begin{pmatrix} -R/3 & & \\ & -R/3 & \\ & & 2R/3 \end{pmatrix}. \tag{I.17}$$

In other words, the general solution of the constraints is given by

$$Y = L_y\, \hat{Y}\, L_y{}^{-1} \tag{I.18}$$

with appropriate $SU(3)$ boosts, L_y. On the other hand, it is possible to express the general solution in the form

$$Y_a{}^b = \left(\frac{1}{3}\delta_a{}^b - \frac{\eta_a \bar{\eta}^b}{\bar{\eta}\eta}\right) R, \tag{I.19}$$

where η_a is a complex triplet and $\bar{\eta}^a = \eta_a^*$. Since Y is unchanged by the complex scaling $\eta_a \to \lambda \eta_a$, this means that the manifold parametrized by L_y can be identified with the 2-dimensional complex projective space, $\mathbb{C}P_2$.

As a supplementary remark we note that it is possible to define the action of a non-compact group on a compact manifold. A trivial example of this is provided by the group of displacements on the real line. By factoring out a discrete subgroup corresponding to displacements which are integer multiples of some unit, $2\pi R$, one obtains a compact quotient space. Formally, one writes

$$L_y = e^{yQ}, \qquad 0 \leqslant y \leqslant 2\pi R, \tag{I.20}$$

where Q is the generator of infinitesimal displacements. Applying an arbitrary displacement, $g = \exp \omega Q$, on the left, one obtains as usual,

$$g L_y = L_{y'} h,$$

where

$$\begin{aligned} y' &= y + \omega - 2\pi R n, \\ h &= (e^{2\pi R Q})^n, \end{aligned} \tag{I.21}$$

and the integer n is determined such that y' lies in the interval $(0, 2\pi R)$. If the points $y = 0$ and $y = 2\pi R$ are identified then the quotient space is a circle of radius R.

A less trivial example is the realization of the non-compact $SO(1, 2)$ by its action on the projective cone, a compact 1-dimensional manifold. That such a realization should exist is easily seen by embedding the cone in 3-dimensional Minkowski space. The null cone

$$Y_0^2 - Y_1^2 - Y_2^2 = 0 \tag{I.22}$$

is clearly invariant under the action of $SO(1, 2)$. But it is a 2-dimensional space. However, the cone is also invariant under the scaling transformation

$$Y_a \to \lambda Y_a. \tag{I.23}$$

A 1-dimensional manifold is obtained by identifying those points which are related by scaling. (In practice one would be dealing with functions on the cone which are homogeneous of degree zero.) A possible parametrization of the manifold would involve writing

$$Y_a = \begin{pmatrix} 1 \\ \sin y \\ \cos y \end{pmatrix}. \tag{I.24}$$

To find the mapping $y \to y'$ corresponding to an hyperbolic rotation in the 02-plane, for example, one would write

$$Y'_a = \begin{pmatrix} ch\omega + sh\omega \cos y \\ \sin y \\ sh\omega + ch\omega \cos y \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \sin y' \\ \lambda \cos y' \end{pmatrix}$$

to obtain the relation

$$\sin y' = \frac{\sin y}{ch\omega + sh\omega \cos y}. \tag{I.25}$$

This kind of realization can be formalized in the manner used previously. The generators, $Q_{ab} = -Q_{ba}$, of $SO(1, 2)$ satisfy the commutation rules

$$[Q_{ab}, Q_{cd}] = \eta_{bc} Q_{ad} - \eta_{ac} Q_{bd} + \eta_{ad} Q_{bc} - \eta_{bd} Q_{ac}. \tag{I.26}$$

It is possible to pick out two of these,

$$Q_{02} \quad \text{and} \quad Q_+ = Q_{01} + Q_{12} \tag{I.27}$$

which are to be identified as the generators of the (non-compact) stability group, H. They satisfy

$$[Q_{02}, Q_+] = Q_+. \tag{I.28}$$

The boosts are defined by

$$L_y = e^{yQ_{12}}. \tag{I.29}$$

Since any element of $SO(1, 2)$ can be represented unambiguously in the form

$$e^{\alpha Q_{12}} e^{\beta Q_+} e^{\gamma Q_{02}} \tag{I.30}$$

it will always be possible to solve the mapping relation

$$gL_y = L_{y'}h$$

for y' and h as functions of y and g.

Another example is the flat group of Scherk and Schwarz [13]. In its simplest version, the generator Q_1 is distinguished from the others, Q_j, $j = 2, 3, \ldots, K$. The commutation rules are

$$\begin{aligned} [Q_i, Q_j] &= 0, \\ [Q_i, Q_1] &= M_{ij} Q_j, \end{aligned} \tag{I.31}$$

where M is antisymmetric, i.e., Q_1 generates orthogonal transformations while the Q_i generate translations. (For $K = 3$ this is simply the 2-dimensional Poincaré algebra.) To construct a compact manifold for the group to act on, consider the elements

$$L_y = e^{y^1 Q_1} e^{y^j Q_j}, \tag{I.32}$$

where y^1 ranges from 0 to $2\pi/m_1$ (m_1 = smallest non-zero eigenvalue of iM_{jk}) and the y^j are restricted to the ranges

$$0 \leqslant y^j \leqslant 2\pi R^j, \qquad j = 2, \ldots, N. \tag{I.33}$$

The action of the group on L_y is easily found. Firstly, the action of $\exp \omega^1 Q_1$ is given by

$$e^{\omega^1 Q_1} L_y = L_{y'}, \tag{I.34}$$

where $y'^j = y^j$ and $y'^1 = y^1 + \omega^1 \pmod{2\pi/m_1}$. Secondly,

$$e^{\omega^j Q_j} L_y = L_{y'} h, \tag{I.35}$$

where $y'^1 = y^1$ and

$$y'^j = y^j + (e^{-y^1 M})_{jk} \omega^k - 2\pi R^j n^j \tag{I.36}$$

with the integers n^j chosen such that y'^j falls in the interval $(0, 2\pi R^j)$. The little group transformation associated with this is then

$$h = \prod_j (e^{2\pi R^j Q_j})^{n_j}. \tag{I.37}$$

The covariant basis is defined as usual by the 1-form

$$\begin{aligned} e &= L_y^{-1} dL_y \\ &= dy^1 (Q_1 - y^j M_{jk} Q_k) + dy^j Q_j, \end{aligned}$$

i.e.,

$$e_\mu{}^\alpha = \begin{matrix} y^1 \\ y^j \end{matrix} \begin{pmatrix} \overset{Q_1}{1} & \overset{Q_k}{-y^i M_{ik}} \\ 0 & \delta_{jk} \end{pmatrix}. \tag{I.38}$$

To construct Killing vectors the matrices $D_\alpha{}^\beta$ of the adjoint representation are needed. These are given by

$$\begin{aligned} D_1{}^\alpha(L_y)\, Q_\alpha &= L_y^{-1} Q_1 L_y \\ &= Q_1 - y^i M_{ik} Q_k, \\ D_j{}^\alpha(L_y)\, Q_\alpha &= L_y^{-1} Q_j L_y \\ &= (e^{y^1 M})_{jk}\, Q_k, \end{aligned}$$

i.e.,

$$D_\alpha{}^\beta(L_y) = \begin{pmatrix} 1 & -y^i M_{ik} \\ 0 & (\exp y^1 M)_{jk} \end{pmatrix}. \tag{I.39}$$

The Killing vectors for left translations are then

$$\begin{aligned} K_\alpha{}^\mu(y) &= D_\alpha{}^\beta(L_y)\, e_\beta{}^\mu \\ &= \begin{pmatrix} 1 & 0 \\ 0 & (\exp y^1 M)_{jk} \end{pmatrix}. \end{aligned} \tag{I.40}$$

(This matrix was denoted $U(y)$ in the work of Scherk and Schwarz.)

We close with the observation that the zero-torsion spin connection $B_\alpha{}^\beta$ (see Appendix 2) is given for this manifold by

$$B_1{}^j = -B_j{}^1 = 0, \qquad B_j{}^k = -dy^1 M_{jk}, \tag{I.41}$$

and the curvature 2-form vanishes, i.e., the manifold is indeed flat.

APPENDIX II: CURVATURE AND TORSION ON G/H

Because of their high degree of symmetry, quotient spaces have rather simple geometrical properties and it is possible to derive a number of explicit formulae to describe these. In this appendix we give derivations for such of these formulae as have been referred to in the text. All of them arise out of manipulations of the boost elements L_y which we use to characterize the manifold G/H.

As has been mentioned in Section III, the first and most important step is to construct the vectors of a covariant basis on the manifold. To this end one must consider the 1-form

$$\begin{aligned} e(y) &= L_y^{-1} dL_y \\ &= e^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}} \\ &= dy^{\mu} e_{\mu}{}^{\hat{\alpha}}(y)\, Q_{\hat{\alpha}}, \end{aligned} \tag{II.1}$$

where the infinitesimal generators of G are denoted $Q_{\hat{\alpha}}$. These generators satisfy the commutation rules

$$[Q_{\hat{\alpha}}, Q_{\beta}] = c_{\hat{\alpha}\beta}{}^{\hat{\gamma}} Q_{\hat{\gamma}} \tag{II.2}$$

with structure constants $c^{\hat{\alpha}\beta\hat{\gamma}}$.

The differential properties of $e(y)$ needed for the discussion of curvature and torsion are summed up in the 2-form

$$\begin{aligned} de(y) &= -dL_y^{-1} \wedge dL_y \\ &= L_y^{-1} dL_y L_y^{-1} \wedge dL_y \\ &= e(y) \wedge e(y), \end{aligned} \tag{II.3}$$

the Cartan–Maurer formula. On substituting the expression $e = e^{\hat{\alpha}} Q_{\hat{\alpha}}$ and using the commutation rules satisfied by the generators, one can extract the equivalent form

$$de^{\hat{\alpha}} = \tfrac{1}{2} e^{\beta} \wedge e^{\hat{\gamma}} c_{\beta\hat{\gamma}}{}^{\hat{\alpha}} \tag{II.4}$$

or, in terms of components,

$$\partial_{\mu} e_{\nu}{}^{\hat{\alpha}} - \partial_{\nu} e_{\mu}{}^{\hat{\alpha}} = -e_{\mu}{}^{\beta} e_{\nu}{}^{\hat{\gamma}} c_{\beta\hat{\gamma}}{}^{\hat{\alpha}}. \tag{II.5}$$

As explained in the text the generators $Q_{\hat{\alpha}}$ separate naturally into two sets, $Q_{\bar{\alpha}}$ belonging to the subgroup H, and Q_{α}, the remainder. The coefficients $e_{\mu}{}^{\alpha}$, $\alpha = 1, 2, \ldots, N$, constitute a set of N covariant basis vectors. The reciprocal basis is denoted $e_{\alpha}{}^{\mu}$, i.e.,

$$e_{\alpha}{}^{\mu} e_{\mu}{}^{\beta} = \delta_{\alpha}{}^{\beta}, \qquad e_{\mu}{}^{\alpha} e_{\alpha}{}^{\nu} = \delta_{\mu}{}^{\nu}. \tag{II.6}$$

The torsion 2-form, T^α, is defined by

$$T^\alpha = de^\alpha + e^\beta \wedge B_\beta{}^\alpha, \tag{II.7}$$

where the 1-form $B_\beta{}^\alpha$ represents the spin connection. Once the latter quantity has been assigned it is possible to construct, in addition to T^α, the curvature 2-form $R_\alpha{}^\beta$ defined by

$$R_\alpha{}^\beta = dB_\alpha{}^\beta + B_\alpha{}^\gamma \wedge B_\gamma{}^\beta. \tag{II.8}$$

One of the simplest possibilities is to choose B such that the torsion vanishes,

$$0 = de^\alpha + e^\beta \wedge B_\beta{}^\alpha. \tag{II.9}$$

On comparing this equation with the Cartan–Maurer formula and noting that $c_{\bar\beta\bar\gamma}{}^\alpha = 0$ (because H is a group) one finds immediately

$$\begin{aligned} B_\beta{}^\alpha &= e^\gamma \tfrac{1}{2} c_{\gamma\beta}{}^\alpha + e^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha \\ &= e^\gamma (\tfrac{1}{2} c_{\gamma\beta}{}^\alpha + \pi_\gamma{}^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha). \end{aligned} \tag{II.10}$$

In this last equation we have used

$$e^{\bar\gamma} = e^\gamma \pi_\gamma{}^{\bar\gamma} = e^\gamma e_\gamma{}^\mu e_\mu{}^{\bar\gamma}. \tag{II.11}$$

From the expression (II.10) for B one obtains the curvature 2-form

$$R_\beta{}^\gamma = \tfrac{1}{2} e^\alpha \wedge e^\epsilon (\tfrac{1}{2} c_{\alpha\epsilon}{}^\delta c_{\delta\beta}{}^\gamma + c_{\alpha\epsilon}{}^{\bar\delta} c_{\bar\delta\beta}{}^\gamma + \tfrac{1}{2} c_{\alpha\beta}{}^\delta c_{\epsilon\delta}{}^\gamma). \tag{II.12}$$

(In deriving this we have used $c_{\alpha\beta}{}^{\bar\gamma} = 0$ which is not generally true. See for example, the discussion of $SO(2, 1)$ in Appendix I. In the general case one would have to keep such terms.)

Another simple possibility is to choose B such that the torsion takes the form

$$T^\alpha = \tfrac{1}{2} e^\beta \wedge e^\gamma c_{\beta\gamma}{}^\alpha. \tag{II.13}$$

Comparison with the Cartan–Maurer formula this time gives

$$B_\beta{}^\alpha = e^{\bar\gamma} c_{\bar\gamma\beta}{}^\alpha \tag{II.14}$$

and hence the curvature

$$R_\beta{}^\gamma = \tfrac{1}{2} e^\alpha \wedge e^\epsilon c_{\alpha\epsilon}{}^{\bar\delta} c_{\bar\delta\beta}{}^\gamma. \tag{II.15}$$

In particular, if the manifold is G itself, i.e., $H = 1$, then this assignment gives vanishing curvature. (On the other hand, for some interesting manifolds such as $SO(N+1)/SO(N)$, the structure constants $c_{\alpha\beta}{}^\gamma$ vanish and the two assignments are identical.)

APPENDIX III: TRANSFORMATION PROPERTIES

Some symmetry properties of the manifold G/H are needed for an understanding of the transformation behaviour of the Yang–Mills fields. The purpose of this appendix is to discuss the derivation of the necessary formula.

Consider firstly the transformation rule (3.8) for the 1-form $e^{\hat{\alpha}}$. For an x-dependent left translation, $y \to y'(x, y)$, this reads

$$e^{\hat{\alpha}}(y') = e^{\beta}(y)\, D_{\beta}{}^{\hat{\alpha}}(h^{-1})^{\hat{\alpha}} + (g^{-1}dg)^{\beta}\, D_{\beta}{}^{\hat{\alpha}}(L_y h^{-1}), \tag{III.1}$$

where $D_{\hat{\alpha}}{}^{\beta}$ is a matrix of the adjoint representation, i.e.,

$$g^{-1}Q_{\hat{\alpha}}\, g = D_{\hat{\alpha}}{}^{\beta}(g)\, Q_{\beta}.$$

The components $(g^{-1}dg)^{\hat{\alpha}}$ and $(hdh^{-1})^{\hat{\alpha}}$ are defined by

$$\begin{aligned} g^{-1}dg &= (g^{-1}dg)^{\hat{\alpha}}\, Q_{\hat{\alpha}}, \\ hdh^{-1} &= (hdh^{-1})^{\hat{\alpha}}\, Q_{\hat{\alpha}} \\ &= (hdh^{-1})^{\bar{\alpha}}\, Q_{\bar{\alpha}}, \end{aligned} \tag{III.2}$$

which is possible since $g^{-1}dg$ belongs to the algebra of G while hdh^{-1} belongs to the subalgebra associated with H. In the following we shall assume that the algebra is fully reducible, i.e.,

$$D_{\alpha}{}^{\bar{\beta}}(h) = D_{\bar{\alpha}}{}^{\beta}(h) = 0. \tag{III.3}$$

This means that (III.1) separates into components from G/H and H, respectively,

$$e^{\alpha}(y') = e^{\beta}(y)\, D_{\beta}{}^{\alpha}(h^{-1}) + (g^{-1}dg)^{\beta}\, D_{\beta}{}^{\alpha}(L_y h^{-1}), \tag{III.4}$$

$$e^{\bar{\alpha}}(y') = e^{\bar{\beta}}(y)\, D_{\bar{\beta}}{}^{\bar{\alpha}}(h^{-1}) + (hdh^{-1})^{\bar{\alpha}} + (g^{-1}dg)^{\beta}\, D_{\beta}{}^{\bar{\alpha}}(L_y h^{-1}). \tag{III.5}$$

Under the transformation $y \to y'(x, y)$ we have

$$dy'^{\mu} = dy^{\nu}\frac{\partial y'^{\mu}}{\partial y^{\nu}} + dx^{m}\frac{\partial y'^{\mu}}{\partial x^{m}}$$

or, equivalently,

$$dy^{\mu} = dy'^{\nu}\frac{\partial y^{\mu}}{\partial y'^{\nu}} + dx^{m}\frac{\partial y^{\mu}}{\partial x^{m}}.$$

Partial derivatives transform such that

$$\begin{aligned} dy^{\mu}\partial_{\mu} + dx^{m}\partial_{m} &= dy'^{\mu}\left(\frac{\partial y^{\nu}}{\partial y'^{\mu}}\partial_{\nu}\right) + dx^{m}\left(\partial_{m} + \frac{\partial y^{\nu}}{\partial x^{m}}\partial_{\nu}\right) \\ &= dy'^{\mu}\partial'_{\mu} + dx^{m}\partial'_{m} \end{aligned}$$

(where ∂'_{m} denotes differentiation with respect to x^{m} at fixed y'^{μ}).

Extracting the coefficients of dy'^{μ} and dx^{m} from the formulae (III.4) and (III.5) gives

$$e_{\mu}{}^{\alpha}(y') = \frac{\partial y^{\nu}}{\partial y'^{\mu}} e_{\nu}{}^{\beta}(y)\, D_{\beta}{}^{\alpha}(h^{-1}), \tag{III.6}$$

$$0 = \frac{\partial y^{\nu}}{\partial x^{m}} e_{\nu}{}^{\beta}(y)\, D_{\beta}{}^{\alpha}(h^{-1}) + (g^{-1}\partial_{m} g)^{\beta}\, D_{\beta}{}^{\alpha}(L_{y} h^{-1}), \tag{III.7}$$

$$e_{\mu}{}^{\bar{\alpha}}(y') = \frac{\partial y^{\nu}}{\partial y'^{\mu}} (e_{\nu}{}^{\beta}(y)\, D_{\beta}{}^{\bar{\alpha}}(h^{-1}) + (h\partial_{\nu} h^{-1})^{\bar{\alpha}}), \tag{III.8}$$

$$0 = \frac{\partial y^{\nu}}{\partial x^{m}} e_{\nu}{}^{\beta}(y)\, D_{\beta}{}^{\bar{\alpha}}(h^{-1}) + (h\partial'_{m} h^{-1})^{\bar{\alpha}} + (g^{-1}\partial_{m} g)^{\beta}\, D_{\beta}{}^{\bar{\alpha}}(L_{y} h^{-1}). \tag{III.9}$$

The formula (III.7) can be arranged in the form

$$\frac{\partial y^{\mu}}{\partial x^{m}} = -(g^{-1}\partial_{m} g)^{\beta}\, K_{\beta}{}^{\mu}(y), \tag{III.10}$$

where $K_{\beta}{}^{\mu}$, the so-called Killing vector, is defined by the expression

$$K_{\beta}{}^{\mu}(y) = D_{\beta}{}^{\gamma}(L_{y})\, e_{\gamma}{}^{\mu}(y). \tag{III.11}$$

Its transformation behaviour is obtained with the help of (III.6), viz.

$$\begin{aligned} K_{\beta}{}^{\mu}(y') &= D_{\beta}{}^{\gamma}(L_{y'})\, e_{\gamma}{}^{\mu}(y') \\ &= D_{\beta}{}^{\gamma}(gL_{y}h^{-1})\, D_{\gamma}{}^{\alpha}(h)\, e_{\alpha}{}^{\nu}(y) \frac{\partial y'^{\mu}}{\partial y^{\nu}} \\ &= D_{\beta}{}^{\hat{\gamma}}(g)\, K_{\hat{\gamma}}{}^{\nu}((y) \frac{\partial y'^{\mu}}{\partial y^{\nu}}. \end{aligned} \tag{III.12}$$

From (III.10) and (III.12) it is a simple exercise to derive an expression for $\partial y'^{\mu}/\partial x^{m}$. Thus,

$$\begin{aligned} \frac{\partial y'^{\mu}}{\partial x^{m}} &= -\frac{\partial y^{\nu}}{\partial x^{m}} \frac{\partial y'^{\mu}}{\partial y^{\nu}} \\ &= (g^{-1}\partial_{m} g)^{\beta}\, K_{\beta}{}^{\nu}(y) \frac{\partial y'^{\mu}}{\partial y^{\nu}} \\ &= (g^{-1}\partial_{m} g)^{\beta}\, D_{\beta}{}^{\hat{\gamma}}(g^{-1})\, K_{\hat{\gamma}}{}^{\mu}(y') \\ &= -(g\partial_{m} g^{-1})^{\beta}\, K_{\beta}{}^{\mu}(y'). \end{aligned} \tag{III.13}$$

With the help of (III.12) and (III.13) one can easily extract the transformation rule for the Yang–Mills gauge fields $A_a{}^{\beta}(x)$ defined by (4.1), or

$$E_a{}^{\mu}(x, y) = A_a{}^{\beta}(x)\, K_{\beta}{}^{\mu}(y). \tag{III.14}$$

Since $E_a{}^{\mu}$ is part of a contravariant vector it must transform according to

$$E_a{}^{\mu}(x, y) \to E'^{\mu}_a(x, y') = E_a{}^{\nu}(x, y)\frac{\partial y'^{\mu}}{\partial y^{\nu}} + E_a{}^{n}(x, y)\frac{\partial y'^{\mu}}{\partial x^{n}}. \tag{III.15}$$

Substitution of (III.14) into this rule gives

$$\begin{aligned} A'^{\beta}_a(x)\, K_{\beta}{}^{\mu}(y') &= A_a{}^{\beta}(x)\, K_{\beta}{}^{\nu}(y)\frac{\partial y'^{\mu}}{\partial y^{\nu}} + E_a{}^{n}(x)\frac{\partial y'^{\mu}}{\partial x^{n}} \\ &= A_a{}^{\beta}(x)\, D_{\beta}{}^{\dot{\gamma}}(g^{-1})\, K_{\dot{\gamma}}{}^{\mu}(y') - E_a{}^{n}(x)(g\partial_n g^{-1})^{\beta}\, K_{\beta}{}^{\mu}(y'). \end{aligned}$$

Since $K_{\beta}{}^{\mu}(y')$ is a common factor in this equation, it may be removed. One is left with the rule

$$A'^{\beta}_a(x) = A_a{}^{\dot{\gamma}}(x)\, D_{\dot{\gamma}}{}^{\beta}(g^{-1}) - E_a{}^{n}(x)(g\partial_n g^{-1})^{\beta}$$

or, equivalently,

$$A'_m(x) = gA_m(x)\, g^{-1} - g\partial_m g^{-1}, \tag{III.16}$$

where $A_m(x) = E_m{}^{a}(x)\, A_a{}^{\beta}(x)\, Q_{\beta}$.

A more direct construction of the Killing vectors makes reference to infinitesimal transformations. Thus, the infinitesimal form of the transformation rule $L_{y'} = gL_y h^{-1}$ reads

$$L_{y+\delta y} = (1 + \delta g^{\dot{\alpha}} Q_{\dot{\alpha}})\, L_y (1 - \delta h^{\bar{\alpha}} Q_{\bar{\alpha}}),$$

i.e.,

$$\delta y^{\mu}\, \partial_{\mu} L_y = \delta g^{\dot{\alpha}} Q_{\dot{\alpha}} L_y - L_y \delta h^{\bar{\alpha}} Q_{\bar{\alpha}}$$

to first order in the infinitesimals δy^{μ}, $\delta g^{\dot{\alpha}}$ and $\delta h^{\bar{\alpha}}$. Multiply on the left by $L_y{}^{-1}$ and use the definition of e,

$$\begin{aligned} \delta y^{\mu} e_{\mu}{}^{\dot{\alpha}} Q_{\dot{\alpha}} &= \delta g^{\dot{\alpha}} L_y{}^{-1} Q_{\dot{\alpha}} L_y - \delta h^{\bar{\alpha}} Q_{\bar{\alpha}} \\ &= \delta g^{\dot{\alpha}} D_{\dot{\alpha}}{}^{\beta}(L_y)\, Q_{\beta} - \delta h^{\bar{\alpha}} Q_{\bar{\alpha}}. \end{aligned}$$

Extract the coefficient of Q_{α}

$$\delta y^{\mu} e_{\mu}{}^{\alpha} = \delta g^{\beta} D_{\beta}{}^{\alpha}(L_y)$$

since $\delta h^{\alpha} = 0$. Hence the infinitesimal δy^{μ} is given by

$$\begin{aligned}\delta y^{\mu} &= \delta g^{\beta} D_{\beta}{}^{\alpha}(L_y)\, e_{\alpha}{}^{\mu}(y) \\ &= \delta g^{\beta} K_{\beta}{}^{\mu}(y)\end{aligned} \tag{III.17}$$

thereby defining the Killing vectors associated with left translations on G/H.

The same procedure could be followed for constructing the Killing vectors associated with right translations. One should use the 1-form $\tilde{e} = LdL^{-1}$ instead of e to obtain $\tilde{K}_{\dot{\alpha}}{}^{\mu} = D_{\dot{\alpha}}{}^{\dot{\beta}}(L_y{}^{-1})\, \tilde{e}_{\dot{\beta}}{}^{\mu}$. For the special case, $H = 1$, the result is particularly simple. One finds $\tilde{K}_{\dot{\alpha}}{}^{\mu} = -e_{\dot{\alpha}}{}^{\mu}$.

APPENDIX IV: HARMONIC EXPANSIONS

The general form of harmonic expansions on the manifold G/H was discussed in Section III. In particular, for the functions $\psi_i(x, y)$ which transform under the combined left translations, $y \to y'$, and associated tangent space rotations, h, such that

$$\psi_i(x, y) \to \psi_i'(x, y') = \mathbb{D}_{ij}(h)\, \psi_j(x, y) \tag{IV.1}$$

the appropriate expansion would be

$$\psi_i(x, y) = \sum_n \sqrt{\frac{d_n}{d_{\mathbb{D}}}} \sum_{\iota, p} D^n_{i\iota, p}((L_y{}^{-1})\, \psi^n_{p\iota}(x). \tag{IV.2}$$

The sum includes all irreducible representations, D^n, of G which contain $\mathbb{D}_{ij}$ on restriction to the subgroup, H. If $\mathbb{D}_{ij}$ occurs more than once in D^n, then the supplementary label ζ is needed. The matrices D^n_{qp} have dimension d_n. The expansion coefficients ψ^n are fields on 4-dimensional spacetime which belong to the irreducible representation D^n,

$$\psi^n_{p\zeta}(x) \to \psi'^n_{p\zeta}(x) = D^n_{pq}(g)\, \psi^n_{q\zeta}(x). \tag{IV.3}$$

They can be projected from $\psi_i(x, y)$ by integrating over the manifold

$$\psi^n_{p\zeta}(x) = \frac{1}{V_K} \sqrt{\frac{d_n}{d_{\mathbb{D}}}} \int_{G/H} d\mu D^n_{p,i\zeta}(L_y)\, \psi_i(x, y). \tag{IV.4}$$

Differentiation with respect to y^{μ} of the fields $\psi_i(x, y)$ is equivalent to an algebraic operation on the components $\psi^n(x)$. The simplest covariant derivative that incorporates $\partial/\partial y^{\mu}$ would be

$$\begin{aligned}\nabla_{\alpha}\psi_i(x, y) &= e_{\alpha}{}^{\mu}\, \partial_{\mu}\psi_i - \tfrac{1}{2} B_{\alpha[\beta\gamma]}\, \mathbb{D}_{ij}(\Sigma^{\beta\gamma})\, \psi_j \\ &= e_{\alpha}{}^{\mu}\, \partial_{\mu}\psi_i + e_{\alpha}{}^{\mu} e_{\mu}{}^{\beta} \mathbb{D}_{ij}(Q_{\beta})\, \psi_j.\end{aligned} \tag{IV.5}$$

In this formula we have used the expression

$$B_{\alpha[\beta\gamma]} = e_\alpha{}^\mu e_\mu{}^{\bar{\alpha}} c_{\bar{\alpha}\beta\gamma},$$

which is one of the possible spin connections on G/H, derived in Appendix II.[6] We have also used the formula (3.17) which fixes the embedding of H in the tangent space group, $SO(N)$,

$$-\tfrac{1}{2} c_{\bar{\alpha}\beta\gamma} \mathbb{D}(\Sigma^{\beta\gamma}) = \mathbb{D}(Q_{\bar{\alpha}}).$$

Differentiation of the boost $L_y{}^{-1}$ leads back to the definitions (3.3) of e, viz.

$$\begin{aligned} \partial_\mu L_y{}^{-1} &= -L_y{}^{-1} \partial_\mu L_y L_y{}^{-1} \\ &= -e_\mu{}^\beta Q_\beta L_y{}^{-1}, \end{aligned}$$

and this implies that, in any representation,

$$\partial_\mu D^n(L^{-1}) = -e_\mu{}^\beta D^n(Q_\beta)\, D^n(L_y{}^{-1}). \tag{IV.6}$$

On comparing this with (IV.5) one finds for the covariant derivative of D^n,

$$\nabla_\alpha D^n_{ip}(L^{-1}) = -D^n_{iq}(Q_\alpha)\, D^n_{qp}(L^{-1}). \tag{IV.7}$$

This formula was used in Section IV to reduce some terms from the $4+N$-dimensional Lagrangian to 4-dimensional form.

Finally, the expansion formulae are sketched for two relativity simple cases.

(1) $SU(2)/U(1)$. This is the case of spherical harmonics on the 2-dimensional sphere. With $G = SU(2)$ generated by $Q_{ab} = -Q_{ba}$, $a, b = 1, 2, 3$ and $H = U(1)$ generated by Q_{12} we can take $y^\mu = (\theta, \varphi)$ with

$$L_y = e^{-\varphi Q_{12}} e^{-\theta Q_{31}}. \tag{IV.8}$$

The covariant basis vectors are contained in the 1-form

$$\begin{aligned} e &= e^{\theta Q_{31}} e^{\varphi Q_{12}} d(e^{-\varphi Q_{12}} e^{-\theta Q_{31}}) \\ &= -d\varphi \cos\theta Q_{12} + d\varphi \sin\theta Q_{23} - d\theta Q_{31}. \end{aligned} \tag{IV.9}$$

Expressed in matrix notation, with rows and columns indicated,

$$e_\mu{}^{\dot{\alpha}} = \begin{matrix} \\ \theta \\ \varphi \end{matrix} \begin{matrix} \begin{matrix} 13 & 23 & 12 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\theta & -\cos\theta \end{pmatrix} \end{matrix}. \tag{IV.10}$$

The last column gives the components of $e_\mu{}^{12} (= e_\mu{}^{\bar{\alpha}}$ in the notation of Section III).

[6] This spin connection corresponds to non-vanishing torsion on G/H. In general the spin connection will contain terms additional to this one, including fluctuation terms, for example.

The tangent space group, $O(2)$, in this case coincides with $H = U(1)$ so the embedding problem is trivial. The 2-vector ϕ_α resolves into helicity $\lambda = \pm 1$ combinations,

$$\frac{1}{\sqrt{2}}(\phi_1 \pm i\phi_2),$$

and likewise for higher rank tensors. For a function, $\phi_\lambda(x, y)$ of helicity λ, integer or half-integer, with respect to the tangent space group we have the expansion

$$\phi_\lambda(x, y) = \sum_j \sqrt{2j+1} \sum_m D^j_{\lambda m}(L_y{}^{-1})\, \phi_{m\lambda}{}^j(x), \tag{IV.11}$$

where j takes the values $|\lambda|$, $|\lambda| + 1$,... and D^j denotes the $2j + 1$-dimensional representation of $SU(2)$,

$$\begin{aligned} D^j_{\lambda m}(L_y{}^{-1}) &= D^j_{\lambda m}(e^{\theta Q_{31}} e^{\omega Q_{12}}) \\ &= d^j_{\lambda m}(\theta)\, e^{im\omega}, \end{aligned} \tag{IV.12}$$

in the notation of Wigner.

Generalization to the N-dimensional sphere $SO(N+1)/SO(N)$ is straightforward in principle though, of course, complicated in detail. There is no embedding to be done since $H = SO(N)$ coincides with the tangent space group. Our second example illustrates a case where this is not so.

(2) $SU(3)/SU(2) \times U(1)$. With $G = SU(3)$ generated by the octet of charges, $Q_a{}^b = -Q_b{}^{a\dagger}$ $(Q_a{}^a = 0)$, $a, b = 1, 2, 3$ and H generated by the subset $Q_\alpha{}^\beta$, $\alpha, \beta = 1, 2$ we can parametrize G/H by a complex 2-vector

$$L_y = e^{y^1 Q_1{}^3 + y^2 Q_2{}^3 - \text{h.c.}} \tag{IV.13}$$

This would not be the most convenient choice for practical computations but we do not intend to present more than a sketch here. The question of interest is how to embed $H = SU(2) \times U(1)$ in the tangent space group, $SO(4)$. The general formula for this, (3.17), could be applied but it is simpler in the present case to work *ab initio*.

The triplet of $SU(3)$ decomposes under the subgroup $SU(2) \times U(1)$ such that

$$3 = 2_{1/3} + 1_{-2/3}, \tag{IV.14}$$

i.e., into an $SU(2)$ doublet with $U(1)$ quantum number $1/3$ and a singlet with $-2/3$. For the octet the decomposition reads

$$8 = 1_0 + 3_0 + 2_1 + 2_{-1}. \tag{IV.15}$$

If the octet is real then the doublets 2_1 and 2_{-1} are related by complex conjugation. In particular, the generators $Q_\alpha{}^3$, $Q_3{}^\alpha$ associated with G/H are in the representations 2_1 and 2_{-1}, respectively. According to the general principle expressed in formula

(3.17), the 4-vector of $SO(4) \sim SU(2) \times SU(2)$ must have the $SU(2) \times U(1)$ content of G/H,

$$(2, 2) = 2_1 - 2_{-1}. \tag{IV.16}$$

This means that the embedding must be such that

$$\begin{aligned} (2, 1) &= 2_0, \\ (1, 2) &= 1_1 + 1_{-1}, \end{aligned} \tag{IV.17}$$

for the 2-spinors of $SO(4)$. An explicit construction for the $SU(2) \times U(1)$ decomposition of a 4-vector, ϕ_α, is easily arranged. From the 2×2 matrix

$$\begin{pmatrix} \phi_4 + i\phi_3 & i(\phi_1 - i\phi_2) \\ i(\phi_1 + i\phi_2) & \phi_4 - i\phi_3 \end{pmatrix} \tag{IV.18}$$

select the first column. This belongs to the representation 2_1, and so we define the spinor ϕ_λ with components

$$\begin{aligned} \phi_{1/2} &= \frac{1}{\sqrt{2}}(\phi_4 + i\phi_3), \\ \phi_{-1/2} &= \frac{1}{\sqrt{2}}(i\phi_1 - \phi_2). \end{aligned} \tag{IV.19}$$

Now the representation 2_1 is found in the $SU(3)$ multiplets 8, 10, 27,..., with triality zero. It follows, therefore, that the harmonic expansion of the 4-vector, $\phi_\alpha(x, y)$, must take the form

$$\phi_\lambda(x, y) = \sum_{n=8,10,27,\ldots} \sqrt{\frac{d_n}{2}} \sum_p D^n_{\lambda p}(L_y^{-1})\, \phi_p{}^n(x), \tag{IV.20}$$

where the coefficients $\phi_p{}^n(x)$ are complex. Expansions for tensors of $SO(4)$ can be constructed by straightforward generalization of this procedure. It is interesting to note, however, that spinors of $SO(4)$ cannot be represented in this way. This is clear from the fact that the $SU(2) \times U(1)$ content of the 2-spinors, viz. 2_0, 1_1 and 1_{-1} is not to be found in any $SU(3)$ multiplet. In fact, it is not possible to define spinors on the manifold $SU(3)/SU(2) \times U(1)$.

APPENDIX V: Spectrum of 4 + 1 Kaluza–Klein Theory

Since the ground state of the $K = 1$ theory is flat, it can be obtained as a vacuum solution of the 4 + 1-dimensional Einstein equations without a cosmological term. For $K > 1$ the manifold G/H has constant curvature and it is necessary to introduce non-geometrical (matter) field to act as sources. Because of the simplicity of the

$K = 1$ theory one can easily obtain its excitation spectrum and relate it to an underlying non-compact symmetry, $SO(1, 2)$.

From the fact that the $4 + 1$-dimensional Einstein theory has five degrees of freedom one expects that all massive states belong to multiplets of spin 2. This is easily verified. Write the metric tensor in the form

$$g_{MN} = \eta_{MN} + h_{MN}, \tag{V.1}$$

where $M, N = 0, 1, 2, 3, 4$ and η_{MN} is the flat metric,

$$\eta_{MN} = \text{diag}(+1, -1, -1, -1, -1). \tag{V.2}$$

Treating the components h_{MN} as small quantities, one obtains the second order terms in the Einstein Lagrangian

$$\begin{aligned}\mathscr{L}_2 = {} & \tfrac{1}{4}(h_{MN,L}h_{MN,L} - 2h_{MN,L}h_{ML,N} + 2h_{MN,N}h_{LL,M} - h_{MM,L}h_{NN,L}) \\ & + \tfrac{1}{2}J_{MN}h_{MN},\end{aligned} \tag{V.3}$$

where J_{MN} is an external source. The connections vanish so that simple partial derivatives are indicated, $h_{MN,L} = \partial_L h_{MN}$. The equations of motion derived from (V.3) take the form

$$h_{MN,LL} - h_{ML,NL} - h_{NL,ML} + h_{LL,MN} + \eta_{MN}h_{KL,KL} - \eta_{MN}h_{KK,LL} = J_{MN}. \tag{V.4}$$

These equations are compatible only if J is conserved,

$$J_{MN,N} = 0. \tag{V.5}$$

This is a relic of the 5-dimensional general covariance. The usual way to solve (V.4), subject to the compatibility condition (V.5), is to impose a co-ordinate condition. For example, in the gauge

$$h_{MN,N} = 0 \tag{V.6}$$

they are solved by

$$h_{MN} = \frac{1}{\partial^2}\left[J_{MN} - \frac{1}{3}\left(\eta_{MN} - \frac{\partial_M\partial_N}{\partial^2}\right)J_{LL}\right], \tag{V.7}$$

where ∂^2 represents the 5-dimensional d'Alembertian.

On substituting the solution (V.7) back into the Lagrangian (V.3), discarding total derivatives, one finds that $\mathscr{L}_2$ reduces to

$$\begin{aligned}\mathscr{L}_2 &= \frac{1}{4}J_{MN}h_{MN} \\ &= \frac{1}{4}\left[J_{MN}\frac{1}{\partial^2}J_{MN} - \frac{1}{3}J_{MM}\frac{1}{\partial^2}J_{NN}\right].\end{aligned} \tag{V.8}$$

This expression represents the effective interaction between conserved (i.e., physical) sources due to the exchange of the Kaluza–Klein particles.

Now consider the pole, $\partial^2 = 0$, in (V.8). The residue simplifies on using the conservation of J_{MN}. Firstly, if $\partial_4 \neq 0$ we can choose a frame in which

$$\partial_0 = \partial_4 \qquad \text{and} \qquad \partial_i = 0, \qquad i = 1, 2, 3. \tag{V.9}$$

In this frame we have

$$J_{0M} = J_{4M}, \qquad M = 0, 1, 2, 3, 4 \tag{V.10}$$

so that $J_{MM} = -J_{ii}$ and $J_{MN}^2 = J_{ij}^2$. It follows that, in the neighbourhood of the pole, the effective interaction (V.8) reduces to

$$\mathscr{L}_2 \simeq \frac{1}{4} J_{ij}^t \frac{1}{\partial_0^2 - \partial_4^2} J_{ij}^t, \tag{V.11}$$

where the physically significant part of the source is the traceless 3-dimensional tensor

$$J_{ij}^t = J_{ij} - \tfrac{1}{3}\delta_{ij} J_{kk}. \tag{V.12}$$

This result indicates that the massive states ($\partial_4 \neq 0$) are indeed purely spin 2. Their masses are given by the eigenvalues of $-\partial_4^2 = (n/2\pi R)^2$, $n = 1, 2, \ldots$, where R is the radius of the Kaluza–Klein circle.

On the other hand, if $\partial_4 = 0$ we have the massless states and they can be analysed conveniently by choosing a frame where

$$\partial_0 = \partial_3 \qquad \text{and} \qquad \partial_1 = \partial_2 = \partial_4 = 0. \tag{V.13}$$

In this frame the effective interaction reduces, in the neighbourhood of the pole, to the form

$$\frac{1}{4} \sum_{\lambda=-2}^{2} J_{-\lambda} \frac{1}{\partial_0^2 - \partial_3^2} J_\lambda, \tag{V.14}$$

where λ refers to the $O(2)$ helicity and

$$\begin{aligned} J_2 &= \frac{1}{2}(J_{11} - J_{22}) + iJ_{12}, \\ J_1 &= J_{41} + iJ_{42}, \\ J_0 &= \frac{1}{\sqrt{6}}(J_{11} + J_{22} - 2J_{44}), \\ J_{-1} &= J_{41} - iJ_{42}, \\ J_{-2} &= \frac{1}{2}(J_{11} - J_{22}) - iJ_{12}. \end{aligned} \tag{V.15}$$

The terms in (V.14) correspond to the exchange of graviton, photon and Bands–Dicke scalar.

Turning now to the non-compact symmetry aspect of this theory, we note that the general co-ordinate transformations in 4 + 1-dimensions include, in particular, the set

$$\delta x^m = 0,$$

$$\delta y = \omega^1(x)\frac{\cos My}{M} + \omega^2(x)\frac{\sin My}{M} + \omega^3(x)\frac{1}{M}, \tag{V.16}$$

where $M = 1/2\pi R$. (The finite transformations are discussed in Appendix I. See, in particular (I.25).) The infinitesimal generators are given by

$$Q_1 = -\frac{\cos My}{M}\partial_y, \qquad Q_2 = -\frac{\sin My}{M}\partial_y, \qquad Q_3 = -\frac{1}{M}\partial_y, \tag{V.17}$$

and these are easily seen to generate the algebra of $SO(1, 2)$,

$$[Q_1, Q_2] = -Q_3, \qquad [Q_2, Q_3] = Q_1, \qquad [Q_3, Q_1] = Q_2. \tag{V.18}$$

Moreover, the Casimir invariant vanishes,

$$\eta^{ab}Q_aQ_b = -\left(\frac{\cos My}{M}\partial_y\right)^2 - \left(\frac{\sin My}{M}\partial_y\right)^2 + \left(\frac{1}{M}\partial_y\right)$$
$$= 0.$$

This means that fields defined on the circle belong to an irreducible representation of $SO(1, 2)$. It is infinite dimensional but not unitary, however. Since the mass operator is given by $-\partial_y{}^2 = M^2Q_3^2$ one sees that the spectrum of masses is simply the spectrum of Q_3^2 in the irreducible representation of $SO(1, 2)$.

Generalizations to higher dimensional manifolds are possible. For example, on the sphere S^K, which is invariant under $SO(K + 1)$, it is possible to define the action of $SO(1, K + 1)$. However, to find the excitation spectrum and perhaps relate it to the spectrum of some combination of generators of $SO(1, K + 1)$, it would be necessary to include the excitations of the matter fields wich are needed for the compactification.

ACKNOWLEDGMENT

The authors wish to thank S. Randjbar-Daemi and W. Mecklenburg for many stimulating and helpful discussions.

REFERENCES

1. TH. KALUZA. *Sitz. Preuss. Akad. Wiss. Berlin, Math. Phys. K*1 (1921), 966.
2. O. KLEIN. *Z. Phys.* **37** (1926), 895.
3. J. SCHERK AND J. H. SCHWARZ, *Phys. Lett. B* **57** (1975), 463; E. CREMMER AND J. SCHERK, *Nucl. Phys. B* **103** (1976), 393; **108** (1976), 409.
4. B. DE WITT. "Dynamical Theory of Groups and Fields." Gordon & Breach, New York, 1965.
5. Z. HORVATH. L. PALLA, E. CREMMER, AND J. SCHERK, *Nucl. Phys. B* **127** (1977), 57; L. PALLA, "Proc. 1978 Tokyo Conf. on High Energy Physics," p. 629.
6. Z. HORVATH AND L. PALLA, *Nucl. Phys. B* **142** (1978), 327.
7. P. G. O. FREUND AND M. A. RUBIN, *Phys. Lett. B* **97** (1980), 233.
8. A. S. SCHWARZ, *Comm. Math. Phys.* **56** (1977), 79; V. N. ROMANOV, A. S. SCHWARZ, AND YU. S. TYUPKIN, *Nucl. Phys. B* **130** (1977), 209; V. N. ROMANOV, I. V. FROLOV, AND A. S. SCHWARZ, *Teor. Mat. Fiz.* **37** (1978), 305; A. S. SCHWARZ AND YU. S. TYUPKIN, *Nucl. Phys. B* **187** (1981), 321; R. CASALBUONI, G. DOMOKOS, AND S. KOVESI–DOMOKOS, *Phys. Rev. D* **17** (1978), 2048.
9. J. F. LUCIANI, *Nucl. Phys. B* **135** (1978), 111; E. WITTEN, *Nucl. Phys. B* **186** (1981), 412.
10. R. KERNER, *Ann. Inst. H. Poincaré* **9** (1968), 143; Y. M. CHO AND P. G. O. FREUND, *Phys. Rev. D* **12** (1975), 1711; Y. M. CHO, *J. Math. Phys.* **16** (1975), 2029; Y. M. CHO AND P. S. JANG, *Phys. Rev. D* **12** (1975), 2789; A. TRAUTMAN, *Rep. Math. Phys.* **16** (1975), 2029; K. I. MACRAE, *Phys. Rev. D* **18** (1978), 3737; C. ORZALESI, Multidimensional unified theories, Parma preprint IFPR/TH 63 (1980); W. MECKLENBURG, *Phys. Rev. D* **21** (1980), 2149.
11. K. NOMIZU, *Amer. J. Math.* **76** (1954), 33; S. KOBAYASHI AND K. NOMIZU, "Foundations of Differential Geometry," Vol. II, Wiley–Interscience, New York, 1969.
12. See, for example, ABDUS SALAM AND J. STRATHDEE, *Phys. Rev.* **184** (1969), 1750; **184** (1969), 1760.
13. J. SCHERK AND J. H. SCHWARZ, *Nucl. Phys. B* **153** (1979), 61.

Reprinted with permission from *Nuclear Physics* **B214** (1983) 491–512
© 1983 North-Holland Publishing Company

SPONTANEOUS COMPACTIFICATION IN SIX-DIMENSIONAL EINSTEIN–MAXWELL THEORY

S. RANDJBAR-DAEMI[1]

International Centre for Theoretical Physics, Trieste, Italy

ABDUS SALAM

International Centre for Theoretical Physics, Trieste, Italy,
and
Imperial College, London, England

J. STRATHDEE

International Centre for Theoretical Physics, Trieste, Italy

Received 8 November 1982

A discrete set of solutions to the classical Einstein–Maxwell equations in six-dimensional space-time is considered. These solutions have the form of a product of four-dimensional constant curvature space-time with a 2-sphere. The Maxwell field has support on the 2-sphere where it represents a monopole of magnetic charge, $n = \pm 1, \pm 2, \ldots$. The spectrum of massless and massive states is obtained for the special case of flat 4-space, and the solution is shown to be classically stable. The limiting case where the radius of the 2-sphere becomes small is considered and a dimensionally reduced effective lagrangian for the long range modes is derived. This turns out to be an $SU(2) \times U(1)$ gauge theory with chiral couplings.

1. Introduction

To illustrate the phenomenon of spontaneous compactification in theories of the Kaluza–Klein type, a six-dimensional model was constructed by Horvath, Palla, Cremmer and Scherk [1, 2]. In this model the gravitational field is coupled to an abelian gauge field*. Three independent parameters are involved, the gravitational and gauge coupling constants and a cosmological constant. Among the possible solutions it was shown that there occurs a candidate ground state in which the geometry factorizes into the product of a four-dimensional space–time of constant curvature with a two-dimensional sphere. In this solution the Maxwell field is non-vanishing only on the 2-sphere where it assumes a magnetic monopole configuration. This non-vanishing ground state value of the Maxwell field contributes to the stress-energy tensor on the right-hand side of the six-dimensional Einstein

[1] Address after 1st December 1982: Institute for Theoretical Physics, University of Vienna, Boltzmanngasse 5, Vienna, Austria.

* A class of non-abelian generalizations has recently been obtained [3].

equations and, in effect, generates the curvature of the 2-sphere. In view of the non-zero "magnetic" charge of the proposed vacuum configuration, it was conjectured that the solution would be a stable one. One of our aims in this article is to test the solution for stability against small perturbations.

Our main purpose, however, is to examine in some detail the Kaluza–Klein mechanism which operates in this model. In particular, we shall compute the excitation spectrum for the special case where the four-dimensional spacetime is flat. It will be shown that the spectrum comprises several series of massive states with spins 0, 1 and 2, and ten massless states: the graviton (helicities ± 2) and gauge vectors (helicities ± 1) corresponding to the local symmetry $SU(2) \times U(1)$. Perhaps surprisingly, there are no massless scalars.

A secondary aim is to illustrate the use of harmonic expansions [4] on the 2-sphere in analyzing the spectrum. This is not entirely trivial and, indeed, we shall find, contrary to the usual assumption, that the SU(2) Yang–Mills 4-vector does not lie entirely in the six-dimensional metric, but is partly in the Maxwell 6-vector as well. In addition, we shall find that its coupling to massless fermions is chiral*.

An important feature of this model is the presence of two independent scales. In the vacuum geometry the curvature of the 2-sphere can be varied independently of the four-dimensional spacetime curvature. This enables us to consider the special case where spacetime is flat** and it further enables us to extract in a meaningful way the effective four-dimensional theory of the zero modes by allowing the radius of the 2-sphere to shrink to zero***.

If the four-dimensional curvature is not adjusted to zero, the candidate ground state would have either de Sitter ($\Lambda > 0$) or anti-de Sitter ($\Lambda < 0$) geometry. It has been shown by Sezgin [6], using the approach of Abbot and Deser [7] that the anti-de Sitter solution is stable in that a conserved, non-negative, energy functional can be defined. Here we consider only the flat case because the interpretation of particle spectra is more straightforward.

This paper is arranged as follows. Sect. 2 sets out the model and the possible ground state solutions are reviewed. Sect. 3 is devoted to the fluctuation spectrum. Fields are expanded around their background values and the linearized equations of motion are obtained. These are solved in the presence of external sources in a suitable gauge. By this means the fluctuation Green functions are in effect obtained. The singularities of these Green functions define the spectrum. It is verified that no tachyons or ghosts are contained. For this investigation it is necessary

* A general discussion of the chirality question can be found in ref. [3a].

** This is to be contrasted with the interesting solution of Freund and Rubin [5] to the 11-dimensional supergravity system where the four-dimensional anti-de Sitter space has curvature of comparable magnitude to that of the "internal" 7-space.

***It should be observed, however, that the shrinking of the 2-sphere leads to a regime of strong couplings for the gauge fields. The classical approximation which we employ cannot be reliable in this limit.

to distinguish physical states from gauge modes and we do this by using appropriately conserved sources. Dimensional reduction is considered in sect. 4 where an effective lagrangian for the zero modes is obtained by shrinking the 2-sphere. Fermions are introduced in sect. 5 where it is shown that zero modes occur whose SU(2) quantum numbers are governed by the Chern class of the monopole background.

2. Ground state background

The six-dimensional Einstein–Maxwell theory with cosmological constant is characterized by the action

$$S = -\int \mathrm{d}^6 z \sqrt{-g}\left[\frac{1}{\kappa^2} R + \tfrac{1}{4} F_{MN} F^{MN} + \lambda\right], \tag{2.1}$$

where R denotes the curvature scalar*, and

$$F_{MN} = \partial_M A_N - \partial_N A_M . \tag{2.2}$$

The field equations are

$$R_{MN} - \tfrac{1}{2} g_{MN} R = -\tfrac{1}{2}\kappa^2 (T_{MN} - \lambda g_{MN}), \tag{2.3}$$

$$\frac{1}{\sqrt{-g}} \partial_M (\sqrt{-g} F^{MN}) = 0 , \tag{2.4}$$

where the stress energy tensor takes the usual Maxwell form,

$$T_{MN} = F_{ML} F_N{}^L - \tfrac{1}{4} g_{MN} F^2 . \tag{2.5}$$

The most symmetric solution to the eqs. (2.3) and (2.4) would of course be six-dimensional Minkowski space with vanishing Maxwell field. However, such a solution could hardly be relevant to the description of a four-dimensional world. We shall therefore restrict our attention to the most symmetric solutions with the structure $M^4 \times M^2$, viz. the product of four- and two-dimensional spaces of constant curvature. Thus, we take

$$g_{MN}\, \mathrm{d}z^M\, \mathrm{d}z^N = g_{mn}(x)\, \mathrm{d}x^m\, \mathrm{d}x^n + g_{\mu\nu}(y)\, \mathrm{d}y^\mu\, \mathrm{d}y^\nu , \tag{2.6}$$

$$A_M\, \mathrm{d}z^M = A_\mu(y)\, \mathrm{d}y^\mu , \tag{2.7}$$

where $g_{mn}(x)$ corresponds to de Sitter or anti-de Sitter space and $g_{\mu\nu}(y)$ to a 2-sphere. The 1-form, $A_\mu\, \mathrm{d}y^\mu$ is required to be invariant (up to a gauge transformation) under rotations of the 2-sphere. We believe that physical interpretations are

* Upper case indices take the values 0, 1, 2, 3, 5, 6. In the following, lower case latin indices will take the values 0, 1, 2, 3, while lower case greek indices are restricted to 5, 6. That is, we write $z^M = (x^m, y^\mu)$. Early alphabet letters will be used as frame labels, while the mid-alphabet is used for world indices. We follow the conventions of Deser and van Nieuwenhuizen [8]: The signature is $-+\cdots+$. The Riemann tensor is defined such that $R^K_{LMN} = \partial_M \Gamma^K_{LN} - \partial_N \Gamma^K_{LM} - \cdots$, and the Ricci tensor is $R_{LM} = R^K_{LMK}$.

less vexed in the anti-de Sitter case (having a global time-like Killing vector and positive energy unitary representations) than they are for the de Sitter world and we shall see that it is possible to exclude the de Sitter solution by appropriate restrictions on the parameters κ, λ and e. Moreover, it can be shown that the anti-de Sitter solution is classically stable [6].

It is convenient to use spherical polar coordinates on the 2-sphere,

$$g_{\mu\nu}\,\mathrm{d}y^{\mu}\,\mathrm{d}y^{\nu} = a^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\,, \tag{2.8}$$

where a, the radius is to be determined. It is *not* an integration constant.

The rotation invariant Maxwell field is given by

$$A_{\mu}(y)\,\mathrm{d}y^{\mu} = \frac{n}{2e}(\cos\theta \mp 1)\,\mathrm{d}\varphi\,, \tag{2.9}$$

where n is a positive or negative integer and e is the U(1) gauge coupling constant. This coupling constant does not appear in the lagrangian (2.1), where all fields are neutral, but one should keep in mind the possibility of charged matter fields to which A would couple. An example of this will be dealt with in sect. 5. In effect we are asking that the theory should be invariant under the gauge transformations

$$A_M \to A_M + \frac{1}{e}\partial_M \Lambda\,,$$

where $\exp(i\Lambda)$ should be single-valued. In (2.9) the monopole configuration is necessarily specified on two patches ($0 \leq \theta < \pi$ and $0 < \theta \leq \pi$, $0 \leq \varphi < 2\pi$, respectively). Where these neighbourhoods overlap, say at $\theta = \pi/2$, $0 \leq \varphi \leq 2\pi$, the two representations of A must be related by a single-valued gauge transformation, $\exp(in\varphi)$. Hence the requirement that n be a positive or negative integer. The field strength corresponding to (2.9) is given by the 2-form,

$$F = \mathrm{d}A = -\frac{n}{2ea^2}\,a\,\mathrm{d}\theta \wedge a\sin\theta\,\mathrm{d}\varphi\,, \tag{2.10}$$

where a is the radius of the 2-sphere. It is clearly rotation invariant.

Corresponding to the field strength (2.10) the components of the stress energy tensor (2.5), in an orthonormal basis, are given by

$$\begin{aligned} T_{ab} &= -\frac{n^2}{8e^2a^4}\eta_{ab}\,, \qquad a, b = 0, 1, 2, 3\,, \\ T_{\alpha\beta} &= 0\,, \\ T_{\alpha\beta} &= \frac{n^2}{8e^2a^4}\delta_{\alpha\beta}\,, \qquad \alpha, \beta = 5, 6\,. \end{aligned} \tag{2.11}$$

The non-vanishing parts of the Riemann tensor, in the same basis are,

$$R_{abcd} = \frac{\kappa^2 \Lambda}{6}(\eta_{ac}\eta_{bd} - \eta_{ad}\eta_{bc}),$$
$$R_{\alpha\beta\gamma\delta} = \frac{1}{a^2}(\delta_{\alpha\gamma}\delta_{\beta\delta} - \delta_{\alpha\delta}\delta_{\beta\gamma}), \tag{2.12}$$

where the four-dimensional cosmological constant, Λ, is to be determined along with the radius a, by substituting (2.12) into the Einstein equations (2.3), or the equivalent form

$$R_{AB} = -\frac{\kappa^2}{2}(T_{AB} - \tfrac{1}{4}\eta_{AB}T + \tfrac{1}{2}\eta_{AB}\lambda).$$

One finds two algebraic equations for Λ and a^2,

$$\Lambda = -\frac{n^2}{16e^2a^4} + \frac{\lambda}{2}, \qquad \frac{2}{\kappa^2 a^2} = \frac{3n^2}{16e^2a^4} + \frac{\lambda}{2}. \tag{2.13}$$

They are solved by

$$\frac{1}{a^2} = \frac{16e^2}{3n^2\kappa^2}\left[1 \pm \sqrt{1 - \frac{3n^2}{32}\frac{\kappa^4\lambda}{e^2}}\right], \tag{2.14}$$

$$\Lambda = \tfrac{2}{3}\lambda - \frac{32}{9n^2}\frac{e^2}{\kappa^4}\left[1 \pm \sqrt{1 - \frac{3n^2}{32}\frac{\kappa^4\lambda}{e^2}}\right]. \tag{2.15}$$

If $\lambda < 0$ then, for each value of the monopole charge, n, there is one positive solution for a^2. This solution corresponds to $\Lambda < 0$ (anti-de Sitter world). On the other hand, if $\lambda > 0$, then there are two positive solutions for a^2, provided n^2 is not too large,

$$n^2 \leqslant \frac{32}{3}\frac{e^2}{\kappa^2\lambda}.$$

At the upper end of this range, however, one finds $\Lambda > 0$ (de Sitter world) which should probably be excluded. To have $\Lambda \leqslant 0$ we require

$$n^2 \leqslant 8\frac{e^2}{\kappa^4\lambda}. \tag{2.16}$$

Of particular interest for our purposes is the case of flat 4-space, $\Lambda = 0$. This occurs when the parameters of the six-dimensional theory are adjusted such that

$$e^2 = \tfrac{1}{8}n^2\kappa^4\lambda, \tag{2.17}$$

and the radius then takes the value

$$a^2 = \tfrac{1}{8}n^2\frac{\kappa^2}{e^2}. \tag{2.18}$$

We shall examine the fluctuations on this background in sect. 3.

We conclude by remarking that the four-dimensional theory based on this vacuum will be invariant under the transformations of Poincaré $\times (SU(2) \times U(1))_{\text{local}}$. The local SU(2) corresponds to x-dependent rotations of the 2-sphere with appropriate U(1) transformations in the tangent space as explained in ref. [4]. The local U(1) corresponds to the four-dimensional part of the Maxwell gauge symmetry.

3. Fluctuation spectrum ($\Lambda = 0$)

The classical test for stability would be to compute the total energy carried by fluctuations on the background solution and show that it is positive. Such an analysis has been carried out for $\Lambda \leq 0$ and these solutions are indeed stable, at least for weak fluctuations [6]. Our purpose here is somewhat different in that we would like to see what kinds of states are produced in this theory. In particular, we shall derive the spectrum of masses for the scalar, vector and tensor excitations on the $\Lambda = 0$ background. In the course of this we shall verify that there are no tachyons or negative metric ghosts: the background is therefore at least perturbatively stable.

In order to obtain the spectrum one must substitute into the action functional (2.1) the expressions

$$g_{MN} = \bar{g}_{MN} + \kappa h_{MN}, \qquad A_M = \bar{A}_M + V_M, \tag{3.1}$$

where $\bar{g}_{MN}$, $\bar{A}_M$ denote one of the ground state solutions obtained above, and h_{MN}, V_M represent the fluctuations. Terms linear in h, V will vanish by virtue of the field equations satisfied by $\bar{g}$, $\bar{h}$. Terms bilinear in h, V will govern the propagation of the weak disturbances. These we retain. Higher order terms, representing interactions will be discarded. In an orthonormal basis the bilinear terms reduce to

$$\begin{aligned} S_2 = \int \mathrm{d}^6 z \sqrt{-\bar{g}} \Big[& -\tfrac{1}{4} h_{AB;C} h_{AB;C} + \tfrac{1}{8} h_{AA;C} h_{BB;C} \\ & + \frac{1}{4}\left(\bar{R} + \frac{\kappa^2}{4}\bar{F}^2 + \kappa^2 \lambda\right)(h_{AB}h_{AB} - \tfrac{1}{2} h_{AA} h_{BB}) \\ & - \tfrac{1}{2}\bar{R}_{BC}(h_{AB}h_{AC} - h_{AA}h_{BC}) \\ & - \frac{\kappa^2}{2}\bar{F}_{BD}\bar{F}_{CD}(h_{AB}h_{AC} - \tfrac{1}{2} h_{AA} h_{BC}) \\ & + \frac{1}{2}\left(R_{ABCD} - \frac{\kappa^2}{2}\bar{F}_{AC}\bar{F}_{BD}\right) h_{AB} h_{CD} \\ & - \tfrac{1}{2} V_{A;B} V_{A;B} + \tfrac{1}{2}\bar{R}_{AB} V_A V_B \\ & + \kappa (h_{AB}\bar{F}_{BC} - \tfrac{1}{4} h_{BB}\bar{F}_{AC})(V_{C;A} - V_{A;C}) \\ & + \tfrac{1}{2} T_{AB} h_{AB} + J_A V_A \Big], \end{aligned} \tag{3.2}$$

where external sources T_{AB} and J_A have been included. The semicolon notation indicates covariant derivatives using the background connection. In general, the bilinear part of the action would reflect the general co-ordinate invariance and U(1) gauge invariance of the original action as invariance under the transformations

$$\delta h_{AB} = \frac{1}{\kappa}(\xi_{A;B} + \xi_{B;A}), \qquad \delta V_A = \omega_{;A} - \bar{F}_{AB}\xi_B . \tag{3.3}$$

We have taken account of this by imposing the gauge conditions

$$h_{AB;B} = \tfrac{1}{2}h_{BB;A}, \qquad V_{A;A} = 0 . \tag{3.4}$$

It should now be observed that the linear equations of motion for h, V obtained from (3.2) are compatible only if the sources are suitably constrained, viz.

$$T_{AB;B} = \kappa \bar{F}_{AB} J_B , \qquad J_{A;A} = 0 . \tag{3.5}$$

In a complete theory these conservation laws would follow from the equations of motion of those fields which contribute to the sources. Here we must view them as a necessary characterization of "physical" sources (i.e. on-shell matrix elements of current operators).

The next step is to solve the linear equations for h and V in terms of T and J. The solutions are then substituted into the functional (3.2). The result reduces to

$$I(T, J) = \tfrac{1}{2}\int \mathrm{d}^6 z \sqrt{-\bar{g}}\,[\tfrac{1}{2}T_{AB}h_{AB} + J_A V_A], \tag{3.6}$$

where h and V are functional of J and T. This functional defines the components of the propagator,

$$S_2 = \tfrac{1}{2}\int \mathrm{d}^6 z \sqrt{-\bar{g}}\,\mathrm{d}^6 z' \sqrt{-\bar{g}'}$$

$$\times (T_{AB}(z) J_A(z)) \begin{pmatrix} \frac{1}{4}\langle h_{AB}(z) h_{CD}(z')\rangle & \frac{1}{2}\langle h_{AB}(z) V_C(z')\rangle \\ \frac{1}{2}\langle V_A(z) h_{CD}(z')\rangle & \langle V_A(z) V_C(z')\rangle \end{pmatrix} \begin{pmatrix} T_{CD}(z') \\ J_C(z') \end{pmatrix} . \tag{3.7}$$

In practice, the computation would be rather awkward because of the curved background associated with the 2-sphere. To cope with this we have found it useful to expand all fields and sources in spherical harmonics on the 2-sphere. The invariance of the sphere under SO(3) implies that the propagators will be diagonal in an isospin-like quantum number $l = 0, 1, 2, \ldots$ In particular, poles of the four-dimensional Fourier transform will be labelled by l. We shall obtain in this way several sequences of masses, the excitation spectrum.

The expansion method is discussed in detail in ref. [4]. Here, in order to fix the notation, we summarize the main features. To begin, viewing the 2-sphere as a quotient space, SU(2)/U(1) introduce the SU(2) boosts, $L_{\theta\varphi}$, to parametrize it,

$$L_{\theta\varphi} = \mathrm{e}^{-\varphi Q_3}\,\mathrm{e}^{-\theta Q_2}\,\mathrm{e}^{\varphi Q_3}, \tag{3.8}$$

(excluding* $\theta = \pi$) where $[Q_1, Q_2] = Q_3$, etc. The action of an SU(2) element, g, on this space is defined by

$$gL_{\theta\varphi} = L_{\theta'\varphi'}h\,, \tag{3.9}$$

where $h = \exp(\zeta Q_3)$ depends on θ, φ and g. The 1-form, $L^{-1}\,\mathrm{d}L$, belongs to the algebra of SU(2) and its components comprise a covariant basis, $E^{\pm}$, and the U(1) connection 1-form, e^3,

$$L^{-1}\,\mathrm{d}L = \frac{1}{a}E^{+}\frac{Q_1 - iQ_2}{\sqrt{2}} + \frac{1}{a}E^{-}\frac{Q_1 + iQ_2}{\sqrt{2}} + e^3 Q_3\,, \tag{3.10}$$

where

$$\begin{aligned} E^{\pm} &= \pm\frac{a}{i\sqrt{2}}\,\mathrm{e}^{\mp i\varphi}\,(\mathrm{d}\theta \mp i\sin\theta\,\mathrm{d}\varphi)\,,\\ e^3 &= -\mathrm{d}\varphi(\cos\theta - 1). \end{aligned} \tag{3.11}$$

The frames $E^{\pm}$ are related to those of sect. 2 by

$$E^{\pm} = \frac{1}{\sqrt{2}}(E^5 \pm iE^6)\,. \tag{3.12}$$

From (3.9) and (3.10) one finds that, under the action of SU(2), the frames and connection transform according to

$$\begin{aligned} E^{\pm}(y) &\to E^{\pm}(y') = E^{\pm}(y)\,\mathrm{e}^{\pm i\zeta}\,,\\ e^3(y) &\to e^3(y') = e^3(y') - \mathrm{d}\zeta\,. \end{aligned} \tag{3.13}$$

Defining the covariant differential of $L^{-1}_{\theta\varphi}$,

$$\begin{aligned} \nabla L^{-1} &= \mathrm{d}L^{-1} + e^3 Q_3 L^{-1}\\ &= E^{+}\nabla_{+}L^{-1} + E^{-}\nabla_{-}L^{-1}\,, \end{aligned} \tag{3.14}$$

* Strictly, one should employ two patches. The parametrization (3.8) is suitable for $0 \leq \theta < \pi$, $0 \leq \varphi < 2\pi$. For the other patch, $0 < \theta \leq \pi$, $0 \leq \varphi < 2\pi$ one should write $L_{\theta\varphi} = \mathrm{e}^{-\varphi Q_3}\,\mathrm{e}^{-\theta Q_2}\,\mathrm{e}^{-\varphi Q_3}$. The action of $L_{\theta\varphi}$ can be seen by embedding S^2 in R^3 where spherical polar co-ordinates give the parametrization

$$\begin{pmatrix} x\\ y\\ z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\varphi\\ \sin\theta\sin\varphi\\ \cos\theta \end{pmatrix} = D(L_{\theta\varphi})\begin{pmatrix} 0\\ 0\\ 1 \end{pmatrix},$$

where

$$D(L_{\theta\varphi}) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\varphi & \sin\varphi & 0\\ -\sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1 \end{pmatrix}.$$

one finds, by simply rearranging (3.10), the useful formula

$$\nabla_{\pm} L^{-1} = -\frac{1}{a} \frac{Q_1 \mp iQ_2}{\sqrt{2}} L^{-1} . \tag{3.15}$$

The behaviour (3.13), in conjunction with the identification (3.12) means that rotations of the 2-sphere must be associated with tangent space rotations in the 5-6 plane in order to preserve the form of the background (2.9). Observe that the background Maxwell field is proportional to $e^3(y)$. Therefore, fields to be expanded must first be decomposed into irreducible representations of these SO(2) rotations, labelled by the "iso-helicity", λ. Let $\phi_\lambda(x, \theta, \varphi)$ be a typical one. It can then be shown that the harmonic expansion of ϕ_λ takes the form

$$\phi_\lambda(x, \theta, \varphi) = \sum_{l \geq |\lambda|} \sqrt{2l+1} \sum_m D^l_{\lambda m}(L^{-1}_{\theta\varphi}) \phi^l_{\lambda m}(x) , \tag{3.16}$$

where $D^l_{\lambda m}$ belongs to the $2l+1$-dimensional unitary irreducible representation of SU(2).

By virtue of the formula (3.15) the computation of covariant derivatives reduces to algebraic manipulation*,

$$\begin{aligned} \nabla_{\pm} D^l_{\lambda m}(L^{-1}) &= -\frac{1}{a\sqrt{2}} D^l_{\lambda m}(Q_{1\mp i2} L^{-1}) \\ &= +\frac{i}{a} \sqrt{\tfrac{1}{2}(l \mp \lambda)(l \pm \lambda + 1)}\, D^l_{\lambda \pm 1, m}(L^{-1}) . \end{aligned} \tag{3.17}$$

This feature leads to considerable simplications in the equations of motion.

Among the fields h_{AB} and V_A we have the following SO(2) irreducible pieces:

$$\begin{aligned} h_{++} &= \tfrac{1}{2}(h_{55} - h_{66} - 2ih_{56}) , \qquad \lambda = 2 , \\ h_{a+} &= \sqrt{\tfrac{1}{2}}(h_{a5} - ih_{a6}) , \qquad \lambda = 1 , \\ h_{+-} &= \tfrac{1}{2}(h_{55} + h_{66}) , \qquad \lambda = 0 , \\ V_+ &= \sqrt{\tfrac{1}{2}}(V_5 - iV_6) , \qquad \lambda = 1 , \end{aligned} \tag{3.18}$$

plus h_{ab}, V_a with $\lambda = 0$ and the complex conjugates $h_{--} = h^*_{++}$, etc. Each of these has an expansion of the form (3.16). The sources, T_{AB} and J_A are likewise expanded.

* We are using anti-hermitian generators, Q, with phase conventions [9] such that

$$\langle \lambda | Q_{1-i2} = \frac{1}{i} \sqrt{(l-\lambda)(l+\lambda+1)}\, \langle \lambda + 1 | ,$$

$$\langle \lambda | e^{\zeta Q_3} = e^{-i\lambda\zeta} \langle \lambda | .$$

Under the tangent space U(1) associated with the global action of an element, g, of SU(2) on S^2,

$$\phi'_\lambda(x, \theta', \varphi') = e^{-i\lambda\zeta} \phi_\lambda(x, \theta, \varphi) .$$

This implies invariance of $E^+ \phi_+$, etc., as it should.

The equations of motion for h and V derived from (3.2) are

$$\nabla^2(h_{AB}-\tfrac{1}{2}\eta_{AB}h_{CC})+(\bar{R}_{AC}h_{CB}+\bar{R}_{BC}h_{CA})$$
$$+\kappa\bar{F}_{BC}(\nabla_A V_C-\nabla_C V_A)+\kappa\bar{F}_{AC}(\nabla_B V_C-\nabla_C V_B)-\kappa\eta_{AB}\bar{F}_{CD}\nabla_C V_D=-T_{AB},$$
$$\nabla^2 V_A+R_{AB}V_B+\kappa\nabla_C(h_{AB}\bar{F}_{BC}+\bar{F}_{AB}h_{BC}+\tfrac{1}{2}h_{BB}\bar{F}_{CA})=-J_A. \qquad (3.19)$$

Considerable simplifications have been brought about in these equations by use of the background solution. The only non-vanishing components of $\bar{R}$ and $\bar{F}$ are, in the basis (3.11),

$$\bar{R}_{+-}=\bar{R}_{-+}=-\frac{1}{a^2}, \qquad \bar{F}_{+-}=-\bar{F}_{-+}=\frac{i}{a}\frac{\sqrt{2}}{\kappa}. \qquad (3.20)$$

Introducing harmonic expansions for the various components into (3.19) and using (3.17) to evaluate the covariant derivatives $\nabla_\pm$, one can immediately extract the equations for the harmonic components. These equations are, suppressing the iso-spin labels l, m,

$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)(h_{ab}+\eta_{ab}h_{+-})+\eta_{ab}\frac{\sqrt{l(l+1)}}{a^2}(V_+-V_-)=-T_{ab}+\tfrac{1}{2}\eta_{ab}T_{cc},$$
$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)h_{aa}+\frac{4}{a^2}h_{+-}+2\frac{\sqrt{l(l+1)}}{a^2}(V_+-V_-)=2T_{+-},$$
$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)V_a+\frac{\sqrt{l(l+1)}}{a^2}(h_{a+}-h_{a-})=-J_a,$$
$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)h_{a\pm}\mp i\frac{\sqrt{2}}{a}\partial_a V_\pm\mp\frac{\sqrt{l(l+1)}}{a^2}V_a=-T_{a\pm}, \qquad (3.21)$$
$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)V_\pm\pm i\frac{\sqrt{2}}{a}\partial_b h_{b\pm}\mp\frac{\sqrt{l(l+1)}}{a^2}(h_{+-}-\tfrac{1}{2}h_{bb})=-J_\pm,$$
$$\left(\partial^2-\frac{l(l+1)}{a^2}\right)h_{\pm\pm}=-T_{\pm\pm}.$$

The solutions of these equations are to be substituted into the expression

$$I(T,J)=\tfrac{1}{2}\int d^6z\,\sqrt{-\bar{g}}(\tfrac{1}{2}T_{AB}h_{AB}+J_A V_A)$$
$$=2\pi a^2\int d^4x\sum_{lm}[\tfrac{1}{2}\bar{T}_{ab}h_{ab}+\tfrac{1}{2}\bar{T}_{++}h_{++}+\tfrac{1}{2}\bar{T}_{--}h_{--}$$
$$+\bar{T}_{a+}h_{a+}+\bar{T}_{a-}h_{a-}+\tfrac{1}{4}\bar{T}h+\bar{J}_a V_a+\bar{J}_+V_++\bar{J}_-V_-]. \qquad (3.22)$$

Because of the reality of h_{AB} and V_A and the properties of $D^l_{\lambda m}$ it follows that the

harmonic components are subject to the reality condition [9]

$$\overline{\phi^{l}_{\lambda m}(x)} = (-)^{\lambda - m}\phi^{l}_{-\lambda,-m}(x)\,. \tag{3.23}$$

This ensures the reality of $I(T, J)$.

To solve the set (3.21) and substitute the result into (3.22) is a straightforward though lengthy computation. We illustrate the approach by finding part of the dependence* of I of T_{++}. The contribution of h_{++} can be found quite easily since h_{++} decouples from the rest. Thus, from (3.21)

$$h_{++} = -\left(\partial^2 - \frac{l(l+1)-2}{a^2}\right)^{-1} T_{++}\,.$$

The contribution to I is then

$$\begin{aligned} &2\pi a^2 \int \mathrm{d}^4x \sum_{lm} [\tfrac{1}{2}\bar{T}_{++}h_{++} + \tfrac{1}{2}\bar{T}_{--}h_{--}) \\ &= 2\pi a^2 \int \mathrm{d}^4x \sum_{lm} \left[-\tfrac{1}{2}\bar{T}_{++}\left(\partial^2 - \frac{l(l+1)^{-2}}{a^2}\right)^{-1} T_{++} + \text{h.c.}\right] \\ &= 2\pi a^2 \int \mathrm{d}^4x \sum_{l\geq 2}\sum_{m} \bar{T}_{++}\left(-\partial^2 + \frac{l(l+1)-2}{a^2}\right)^{-1} T_{++}\,. \end{aligned}$$

To interpret this result one can Fourier analyze T_{++} obtaining

$$2\pi a^2 \int \frac{\mathrm{d}^4p}{(2\pi)^4} \sum_{l\geq 2}\sum_{m} \frac{|\tilde{T}_{++}(p)|^2}{p^2 + [l(l+1)-2]/a^2}\,.$$

One sees that it describes the exchange of particles of mass $(l(l+1)-2)/a^2$ carrying quantum numbers appropriate to the vertex factor $\tilde{T}^{l}_{++m}(p)$, i.e. $2l+1$-dimensional multiplets of scalar particles. The overall factor $2\pi a^2$ is not significant, it can be scaled away.

It is important to make use of the conservation laws (3.5) since only the physical part of the spectrum is of interest. With the help of (3.17) one can extract from (3.5) the conservation laws for the harmonic components,

$$\begin{aligned} &\partial_a T_{ab} + \frac{i}{a}\sqrt{\frac{l(l+1)}{2}}\,(T_{b+} + T_{b-}) = 0\,, \\ &\partial_a T_{a\pm} + \frac{i}{2}\sqrt{\frac{l(l+1)}{2}}\,T_{-+} + \frac{i}{a}\sqrt{\frac{(l+2)(l-1)}{2}}\,T_{\pm\pm} = \pm i\frac{\sqrt{2}}{a}J_{\pm}\,, \\ &\partial_a J_a + \frac{i}{a}\sqrt{\frac{l(l+1)}{2}}\,(J_+ + J_-) = 0\,. \end{aligned} \tag{3.24}$$

* The full dependence of I on T_{++} can only be obtained by solving the complete set of equations (3.21). The result is contained I_0, given below, cf. (3.26).

The solution of (3.21) is given in the appendix. Here we exhibit only the pole terms which result when the subsitution into (3.22) is made. The expressions are simplified by referring them to an appropriate Lorentz frame. To see the massive states it is best to use the rest frame, $p_a = (p_0, 0, 0, 0)$. For massless states we use instead the frame $p_a = (p_0, 0, 0, p_3)$. The spin of the exchanged particle is signalled by the source currents appearing in the residue.

Taking first the massive excitations one can show that the pole terms in $I(T, J)$ arrange themselves into six towers,

$$I = \sum_{l\geq 0} (I_0^{(+)} + I_0^{(-)} + I_2) + \sum_{l\geq 1} (I_1^{(+)} + I_1^{(-)}) + \sum_{l\geq 2} I_0 , \tag{3.25}$$

where the subscripts 0, 1, 2 indicate the spin of the exchanged particles which contribute. The explicit forms are

$$\begin{aligned}
I_0 &= \frac{1}{p^2+M_0^2}\,\frac{1}{2l(l+1)}\,|T_{++}-T_{--}|^2 ,\\
I_0^{(\pm)} &= \frac{1}{p^2+M_{0\pm}^2}\,\frac{1}{4\sqrt{1+12l(l+1)}}\Big|\sqrt{\tfrac{4}{8}}[5\pm\sqrt{1+12l(l+1)}]^{1/2}(\tfrac{3}{4}T_{00}+\tfrac{1}{4}T_{kk})\\
&\quad+\Big[\frac{\pm(3l(l+1)-2)\sqrt{1+12l(l+1)}-9l(l+1)-2}{l(l+1)-2}\Big]^{1/2} T_{+-}\\
&\quad+i\Big[\frac{l(l+1)}{l(l+1)-2}\,(5\mp\sqrt{1+12l(l+1)})\Big]^{1/2}(J_+ + J_-)\Big|^2 ,\\
I_1^{(\pm)} &= \frac{1}{p^2+M_{1\pm}^2}\,|T_{k+}-T_{k-}\pm\sqrt{2}J_k|^2 ,\\
I_2 &= \frac{1}{p^2+M_2^2}\,|T_{jk}-\tfrac{1}{3}\delta_{jk}T_{il}|^2 ,
\end{aligned} \tag{3.26}$$

where the non-zero masses are given by

$$\begin{aligned}
M_0^2 &= (l-1)(l+2)/a^2 , \qquad l\geq 2\\
M_{0\pm}^2 &= (2l(l+1)+1\pm\sqrt{1+12l(l+1)})/2a^2 , \qquad l\geq 0\\
M_{1\pm}^2 &= (l(l+1)\pm\sqrt{2l(l+1)})/a^2 , \qquad l\geq 1\\
M_2^2 &= l(l+1)/a^2 , \qquad l\geq 0 .
\end{aligned} \tag{3.27}$$

It must be emphasized that these mass formulae have been derived by using the rest frame so that only the non-zero values belong to the spectrum. It will be observed that $M_{0-}=0$ for $l=0, 1$, $M_{1-}=0$ for $l=1$, and $M_2=0$ for $l=0$. These zeroes are artifacts of the computational procedure. To find the true massless states it is necessary to repeat the computation in a more appropriate frame, $p_a =$

$(p_0, 0, 0, p_3)$. When this is done, one finds the terms

$$I(M=0)=\frac{1}{p^2}\{[\tfrac{1}{4}|T_{11}-T_{22}|^2+|T_{12}|^2+|J_1|^2+|J_2|^2]_{l=0}$$
$$+[|T_{1+}+T_{1-}-\sqrt{2}J_1|^2+|T_{2+}+T_{2-}-\sqrt{2}J_2|^2]_{l=1}\}, \qquad (3.28)$$

indicating a total of ten massless states: the graviton (helicities $\lambda = \pm 2$), a "photon" ($\lambda = \pm 1$), and a Yang–Mills triplet ($\lambda = \pm 1, l = 1$). Note, in particular, the absence of zero-helicity massless states.

Finally, we observe that there are no tachyons and, since the residues all have the same sign, no negative metric states. This means that the ground state is at least perturbatively stable.

4. Zero-mode ansatz

The purpose of this section is to derive an effective four-dimensional lagrangian for the long-range sector of the six-dimensional system. That is, we shall suppose that the radius a tends to zero while discarding from the theory all those harmonic components corresponding to masses $\sim 1/a$. This is not an entirely satisfactory kind of limit: it corresponds, in fact, to letting the Maxwell gauge coupling grow without limit,

$$e=\frac{n}{2\sqrt{2}}\frac{\kappa}{a}\to\infty .$$

Of course, the model cannot pretend to be realistic, and with the limited set of parameters available we cannot avoid this difficulty. Our aim here is only to illustrate the procedure of "dimensional reduction" in the context of a particularly simple model.

One of the features to be brought out is the rather complicated placing of the SU(2) gauge field in the six-dimensional system. As mentioned in sect. 1, the Yang–Mills vector turns out to be a mixture of the six-dimensional metric and Maxwell fields. This is contrary to the common assumption that Yang–Mills fields are purely metrical in origin. We shall argue, however, that such mixing effects are to be expected whenever the "matter" fields which participate in the spontaneous compactification carry the appropriate quantum numbers.

Since we know from the analysis of sect. 3 that no scalar zero-modes are present we do not need to introduce any scalar fields in the zero-mode ansatz. We need to account only for the graviton and the $SU(2)\times U(1)$ vectors. We also know from sect. 3 that the helicity-1 triplet is a mixture of metric and Maxwell fields. In order to take account of this mixing we shall start by introducing two independent, $l = 1$, 4-vectors in the ansatz. One combination will reveal itself as massive and we

shall then discard it. Thus, our starting ansatz, in the notation of 1-forms*, reads

$$
\begin{aligned}
E^a(x,y) &= \mathrm{d}x^m E^a_m(x)\,,\\
E^{\alpha}(x,y) &= a\,\mathrm{d}y^{\mu} e^{\alpha}_{\mu}(y) - \kappa\,\mathrm{d}x^m\, W^{\hat{\alpha}}_m(x) D^{\alpha}_{\hat{\alpha}}(L_y)\\
&= a e^{\alpha}(y) - \kappa W^{\hat{\alpha}}(x) D^{\alpha}_{\hat{\alpha}}(L_y)\,, \qquad (4.1)\\
A(x,y) &= \mathrm{d}x^m V_m(x) + \frac{n}{2e}\left(\mathrm{d}y^{\mu} e^3_{\mu}(y) - \frac{\kappa}{a}\mathrm{d}x^m\, U^{\hat{\alpha}}_m(x) D^3_{\hat{\alpha}}(L_y)\right)\\
&= V(x) + \frac{n}{2e}\left(e^3(y) - \frac{\kappa}{a} U^{\hat{\alpha}}(x) D^3_{\hat{\alpha}}(L_y)\right),
\end{aligned}
$$

where $a = 0, 1, 2, 3$; $\alpha = +, -$; $\hat{\alpha} = 1, 2, 3$ and $y^{\mu} = (\theta, \varphi)$. At this stage the system includes the graviton field, $E^a_m(x)$, a U(1) vector, $V_m(x)$, and two SU(2) vectors, $W^{\hat{\alpha}}_m(x)$ and $U^{\hat{\alpha}}_m(x)$.

The symmetry group of the background includes rotations of the 2-sphere coupled to appropriate frame rotations and U(1) gauge transformations. The frame rotations are determined by the invariance of $e^{\alpha}(y)$ and the U(1) transformations by the invarince of $e^3(y)$. As discussed in sect. 3, a left translation, g, of the quotient space, SU(2)/U(1) (parametrized by L_y), induces a U(1) rotation, h, such that $gL_y = L_y h$. On writing $h = \exp(\zeta Q_3)$ one finds

$$
e^{\pm}(y') = e^{\pm}(y)\exp(\pm i\zeta)\,, \qquad e^3(y') = e^3(y) - \mathrm{d}\zeta\,, \qquad (4.2)
$$

where $\zeta = \zeta(y, g)$. Invariance is obtained if the frames $E^{\pm}$ are rotated through ζ and the Maxwell field A undergoes a gauge transformation with parameter $\Lambda = n\zeta/2$. This much is enough to give invariance under x-independent left-translations. In order to get local invariance it is necessary to consider the behaviour of $W^{\hat{\alpha}}$ and $U^{\hat{\alpha}}$. Although these fields are associated with excitations on the background, their transformations include inhomogeneous terms whose contributions are needed in order to make manifest the background invariance with respect to *local* left-translations on S^2. The local version** of (4.2) is

$$
\begin{aligned}
e^{\alpha}(y') &= e^{\beta}(y) D^{\alpha}_{\beta}(h^{-1}) + (g^{-1}\,\mathrm{d}g)^{\hat{\beta}} D^{\alpha}_{\beta}(L_y h^{-1})\,,\\
e^3(y') &= e^3(y) - \mathrm{d}\zeta + (g^{-1}\,\mathrm{d}g)^{\hat{\beta}} D^3_{\hat{\beta}}(L_y)\,,
\end{aligned} \qquad (4.3)
$$

where $g^{-1}\,\mathrm{d}g = \mathrm{d}x^m\, g^{-1}\partial_m g$ belongs to the algebra of SU(2). Taken in conjunction with the ansatz (4.1) these identities allow one to deduce the transformation properties of V, W, U. The group acts on E^{α} and A such that

$$
\begin{aligned}
E^{\alpha}(x,y) &\to E'^{\alpha}(x,y') = E^{\beta}(x,y) D^{\alpha}_{\beta}(h^{-1})\,,\\
A(x,y) &\to A'(x,y') = A(x,y) - \frac{n}{2e}\mathrm{d}\zeta\,.
\end{aligned} \qquad (4.4)
$$

* The 1-forms $e^{\pm}(y)$ which appear here are related to $E^{\pm}(y)$ defined in eq. (3.11) by $ae^{\pm}(y) = E^{\pm}(y)$.

** Ref. [4], appendix 3.

Using (4.3) it is a simple matter to show that the form of the ansatz is maintained provided

$$U'^{\hat{\alpha}}(x) = U^{\hat{\beta}}(x) D^{\hat{\alpha}}_{\hat{\beta}}(g^{-1}) - \frac{a}{\kappa}(g\,\mathrm{d}g^{-1})^{\hat{\alpha}},$$
$$W'^{\hat{\alpha}}(x) = W^{\hat{\beta}}(x) D^{\hat{\alpha}}_{\hat{\beta}}(g^{-1}) - \frac{a}{\kappa}(g\,\mathrm{d}g^{-1})^{\hat{\alpha}}. \tag{4.5}$$

That is, *both* U and W transform as SU(2) gauge fields. (The U(1) vector, $V(x)$, is invariant.) It is now clear that the difference, $U - W$, transforms homogeneously and could therefore be expected to acquire a mass. This indeed happens. A straightforward computation gives the field strength 2-form, $F = \mathrm{d}A$,

$$\begin{aligned} F = &-\tfrac{1}{2}\,\mathrm{d}x^m \wedge \mathrm{d}x^n \{V_{mn} - \sqrt{2}(W^{\hat{\alpha}}_{mn} + \nabla_m (U-W)^{\hat{\alpha}}_n - \nabla_n (U-W)^{\hat{\alpha}}_m) D^3_{\hat{\alpha}}(L)\} \\ &+ \mathrm{d}x^m \wedge E^\beta \frac{\sqrt{2}}{a} (U-W)^{\hat{\alpha}}_m D_{\hat{\alpha}}{}^{\gamma}(L) \varepsilon_{\beta\gamma 3} \\ &- \tfrac{1}{2} E^\alpha \wedge E^\beta \frac{n}{2ea^2} \varepsilon_{\alpha\beta 3}, \end{aligned} \tag{4.6}$$

where ∇_m denotes the SU(2) covariant derivative

$$\nabla_m (U-W)^{\hat{\alpha}}_m = \partial_m (U-W)^{\hat{\alpha}}_n - \frac{\kappa}{a} W^{\hat{\beta}}_m (U-W)^{\hat{\gamma}}_n \varepsilon_{\hat{\beta}\hat{\gamma}\hat{\alpha}}, \tag{4.7}$$

and W_{mn} the Yang–Mills field strength. When the components, F_{AB}, are substituted into the lagrangian and the integration over y is performed, there results the expression,

$$-\tfrac{1}{4}V^2_{ab} - \tfrac{1}{6}(W_{ab} + \nabla_a (U-W)_b - \nabla_b (U-W)_a)^2 - \frac{1}{6}\frac{4}{a^2}(U_a - W_a)^2 - \frac{n^2}{8e^2a^4}. \tag{4.8}$$

On the other hand, the vector W appears in E^α and so will contribute to the six-dimensional curvature scalar, R_6. The contribution of R_6 to the four-dimensional lagrangian has been discussed in the literature [10]. It gives:

$$-\frac{1}{\kappa^2} R_4 - \tfrac{1}{6} W^2_{ab} + \frac{1}{\kappa^2 a^2}. \tag{4.9}$$

On combining (4.9) with (4.8) and introducing the new combinations,

$$A = \sqrt{\tfrac{1}{3}}(U+W), \qquad X = \sqrt{\tfrac{1}{3}}(U-W), \tag{4.10}$$

one finds,

$$\begin{aligned} &-\frac{1}{\kappa^2} R_4 - \tfrac{1}{4} V^2_{ab} - \frac{1}{8}\left[A_{ab} - (\nabla_a X_b - \nabla_b X_a) - \frac{\kappa}{a}\frac{\sqrt{3}}{2} X_a \times X_b\right]^2 \\ &\quad - \frac{1}{8}\left[A_{ab} + (\nabla_a X_b - \nabla_b X_a) + \frac{3\kappa}{a}\frac{\sqrt{3}}{2} X_a \times X_b\right]^2 - \frac{1}{2}\frac{4}{a^2} X^2_a \\ &= -\frac{1}{\kappa^2} R_4 - \tfrac{1}{4} V^2_{ab} - \tfrac{1}{4} A^2_{ab} - \tfrac{1}{4}(\nabla_a X_b - \nabla_b X_a)^2 - \frac{1}{2}\frac{4}{a^2} X^2_a + \cdots, \end{aligned} \tag{4.11}$$

where A_{ab} denotes the Yang–Mills field strength made out of A and the covariant derivative ∇_a is now defined by

$$\nabla_a X_b = \partial_a X_b - \frac{\sqrt{3}}{2}\frac{\kappa}{a} A_a \times X_b . \tag{4.12}$$

It now appears that A_a is the true Yang–Mills vector, while X_a is massive. (It corresponds to the mass $M_{1+}(l=1)$ of eq. (3.27).) In passing, we note that the SU(2) coupling constant is given by

$$f = \frac{\sqrt{3}}{2}\frac{\kappa}{a} = \frac{\sqrt{6}}{n} e , \tag{4.13}$$

where e is the U(1) coupling.

This concludes our discussion of the bosonic part of the long-range sector of this model. In the next section we consider the introduction of fermions and show how fermionic zero-modes can arise.

5. Fermion spectrum

At the classical level fermion fields do not affect the ground state solution and the question of its stability. However, after establishing the properties of the classical ground state one is free to consider how it affects the propagation of fermionic excitations. In particular, one can demonstrate the presence of zero-mode solutions of the Dirac equation.

Fermion fields should be treated as scalars under the group of general co-ordinate transformations and as spinors under the tangent space group SO(1, 5). The simplest examples are 4-spinors and we shall treat only these. The relevant Clifford algebra is generated by eight-dimensional matrices, Γ_A, which satisfy

$$\begin{aligned} \tfrac{1}{2}\{\Gamma_A, \Gamma_B\} &= \eta_{AB} \\ &= \operatorname{diag}(-1, 1, 1, 1, 1, 1) . \end{aligned} \tag{5.1}$$

(On occasion we shall also use the combinations $\Gamma_5 \mp i\Gamma_6 = \sqrt{2}\Gamma_\pm$ appropriate to the spherical basis on the 2-sphere, i.e. $\eta_{+-} = \eta_{-+} = 1$, $\eta_{++} = \eta_{--} = 0$.) It is convenient to employ a particular realization of the algebra (5.1) which is adapted to the reduction, $SO(1, 5) \to SO(1, 3) \times SO(2)$. In this realization we have

$$\begin{aligned} &\Gamma_a = \gamma_a \times \tau_1 , \\ &\Gamma_5 = \gamma_5 \times \tau_1 , \qquad a = 0, 1, 2, 3 , \\ &\Gamma_6 = 1 \times \tau_2 , \end{aligned} \tag{5.2}$$

where γ_a and γ_5 denote the usual four-dimensional matrices,

$$\tfrac{1}{2}\{\gamma_a, \gamma_b\} = \eta_{ab} , \qquad \gamma_5^2 = 1 . \tag{5.3}$$

The generators of SO(1, 5) are given by

$$-\tfrac{1}{2}\Gamma_{AB} = -\tfrac{1}{4}\{\Gamma_A, \Gamma_B\} \tag{5.4}$$

and they evidently commute with τ_3. The fundamental representations, 4 and $\bar{4}$, correspond to the eigenvalues +1 and −1 or τ_3, respectively.

Let $\psi(x; y)$ belong to the representation 4, i.e.

$$\tau_3\psi = \psi\,. \tag{5.5}$$

From (5.2), (5.4) we see that the SO(2) group of rotations in the 5–6 plane is generated by

$$-\tfrac{1}{2}\Gamma_{56} = -\tfrac{1}{2}i\gamma_5\tau_3\,. \tag{5.6}$$

Acting on the chiral components of ψ it gives

$$-\tfrac{1}{2}\Gamma_{56}\psi_{\mathrm{L}} = \tfrac{1}{2}i\psi_{\mathrm{L}}\,, \qquad -\tfrac{1}{2}\Gamma_{56}\psi_{\mathrm{R}} = -\tfrac{1}{2}i\psi_{\mathrm{R}}\,, \tag{5.7}$$

where ψ_{L} and ψ_{R} correspond to $\gamma_5 = -1$ and $\gamma_5 = +1$, respectively.

Now suppose that $\psi(x, y)$ carries the U(1) charge, e. Under local U(1) transformations defined by

$$A \to A - \frac{1}{e}\mathrm{d}\Lambda \tag{5.8}$$

we have

$$\psi \to \mathrm{e}^{i\Lambda}\psi\,, \tag{5.9}$$

where $\Lambda = \Lambda(x, y)$ is a real scalar. Derivatives of ψ which are covariant with respect to both the frame rotations, SO(1, 5), and the U(1) gauge transformations are contained in the 1-form

$$\nabla\psi = \mathrm{d}\psi + B\psi + ieA\psi\,, \tag{5.10}$$

where B denotes the spin connection 1-form

$$B = \tfrac{1}{2}\,\mathrm{d}z^M B_{M[AB]}(-\tfrac{1}{2}\Gamma_{AB})\,. \tag{5.11}$$

For the purposes of this section, where only fermion bilinears are to be retained, it will be sufficient to use background values for B and A, viz.

$$\bar{B} = -\tfrac{1}{2}\Gamma_{56}e^3\,, \qquad \bar{A} = \frac{n}{2e}e^3\,, \tag{5.12}$$

where e^3 is the 1-form, $-\mathrm{d}\varphi(\cos\theta - 1)$, given in sect. 3. In effect, the background covariant derivative of ψ is given by

$$\begin{aligned}\bar{\nabla}\psi &= \mathrm{d}\psi + \tfrac{1}{2}i(n + i\Gamma_{56})e^3\psi \\ &= \mathrm{d}\psi + i\Big(\frac{1+\gamma_5}{2}\,\frac{n-1}{2} + \frac{1-\gamma_5}{2}\,\frac{n+1}{2}\Big)e^3\psi\,.\end{aligned} \tag{5.13}$$

The invariance group of the ground state involves the simultaneous action of left-translations on the 2-sphere, rotations of the E^5 and E^6 frames and a U(1) gauge transformation. In the notation of (3.13), the tangent space rotation through the angle $\zeta = \zeta(\theta, \varphi, g)$ must be accompanied by a gauge transformation with $\Lambda = +\frac{1}{2}n\zeta$ in order to preserve the form of $\bar{A}$, as discussed in sect. 4. With respect to this background symmetry the effective iso-helicities of ψ_R and ψ_L become $\frac{1}{2}-\frac{1}{2}n$ and $-\frac{1}{2}-\frac{1}{2}n$, respectively, as indicated in (5.13).

Now consider the Dirac operator (using $\tau_3\psi = \psi$),

$$\begin{aligned}\Gamma^A\nabla_A\psi &= (\gamma^a\tau_1\partial_a + \gamma^5\tau_1\nabla_5 + \tau_2\nabla_6)\psi \\ &= \tau_1(\not{\partial} + \gamma_5\nabla_5 + i\nabla_6)\psi \\ &= \tau_1\left(\not{\partial} + \frac{1+\gamma_5}{2}\sqrt{2}\nabla_- - \frac{1-\gamma_5}{2}\sqrt{2}\nabla_+\right)\psi .\end{aligned}$$

It appears that the mass term for the 4-spinor takes the form

$$\sqrt{2}(\bar{\psi}_L\nabla_-\psi_R - \bar{\psi}_R\nabla_+\psi_L) . \tag{5.14}$$

To evaluate the covariant derivatives in this expression we must introduce the harmonic expansions,

$$\psi_{R,L} = \sum_{l\geqslant|\lambda_{R,L}|}\sqrt{2l+1}\sum_m D^l_{\lambda m}(L^{-1}_{\theta\varphi})\psi^l_{R,Lm}(x) , \tag{5.15}$$

where the relevant iso-helicities are

$$\lambda_R = \frac{1-n}{2}, \qquad \lambda_L = -\frac{1+n}{2} . \tag{5.16}$$

In terms of the harmonic components the mass term (5.13) becomes

$$\frac{i}{a}\sum_{lm}\sqrt{(l+\tfrac{1}{2})^2 - \tfrac{1}{4}n^2}(\bar{\psi}_L\psi_R - \bar{\psi}_R\psi_L) . \tag{5.17}$$

There is evidently one massless multiplet with

$$l = -\tfrac{1}{2} + \tfrac{1}{2}|n| .$$

This value of l contributes to ψ_R but not to ψ_L if n is a positive integer. Conversely, if n is negative then ψ_L contains the massless mode*.

The coupling of the massless fermion field to the boson fields V, A and E can be derived by substituting the zero-mode ansatz into the six-dimensional Dirac lagrangian

$$i\bar{\psi}\Gamma^A E^M_A(\partial_M + B_M + ieA_M)\psi , \tag{5.18}$$

* The n zero-modes of $\Gamma_+\nabla_- + \Gamma_-\nabla_+$ associated with the magnetic charge, n, are of course predicted by the Atiyah–Singer index theorem.

and integrating over y. Because of the local SU(2) × U(1) symmetry, the result must be

$$i\bar{\psi}_R\gamma^a E_a^m(\partial_m + \tfrac{1}{4}B_{m[bc]}\gamma^{bc} - fA_m^{\hat{a}}I_{\hat{a}} + ieV_m)\psi_R\,, \tag{5.19}$$

if ψ_R is the massless field ($n > 0$). The fields ψ_R belong to the right-handed doublet of SO(1, 3) and to the n-dimensional representation of SU(2) (with generators denoted $I_{\hat{a}}$). The SO(1, 3) spin connection is denoted $B_{m[bc]}$.

One may add more than one spinor field to the lagrangian: for example, consider two fields ψ, ψ' with U(1) couplings Y and Y'. From the iso-helicity formulae for ψ:

$$\lambda_R = \frac{1-nY}{2}, \qquad \lambda_L = -\left(\frac{1+nY}{2}\right),$$

and with similar formulae for ψ' with Y replaced by Y', we would obtain for the $n = 1$ case, a massless iso-singlet ψ_R for $Y = 1$, and a massless iso-doublet ψ'_L for $Y' = -2$. For $n = 2$, with $Y = 1$, $Y' = -1$, we would obtain two "iso-doublets" ψ_R and ψ'_L, both massless. For $n = 2$, the formulae for the ratio of U(1) and SU(2) couplings, $\tan\theta_0 = \sqrt{\frac{1}{6}}n$ (cf. eq. (4.13)) yields $\sin^2\theta = \frac{2}{5}$ – a value very close to the conventional unrenormalized value $\frac{3}{8}$. A determination of "renormalized" $\sin^2\theta$ with $\sin^2\theta_0 = \frac{2}{5}$ as input in a usual GUT context would give a 3% deviation from the value deduced with $\sin^2\theta_0 = \frac{3}{8}$ as input. This accidental agreement is perhaps devoid of physical significance, since we do not expect ψ_R and ψ'_L to correspond to the leptonic or quark doublets. Even so, the fact that $\sin^2\theta$ can in principle be determined by the n-quantum number (corresponding to the monopole-charge in the internal space) is an important consequence of the Kaluza–Klein picture of the influence of extra dimensions on physics.

6. Conclusion

The model discussed in this article, six-dimensional Einstein–Maxwell theory (with fermions), is not intended as a realistic physical theory. We have analyzed it in some detail because it seems to represent the simplest possible version of a Kaluza–Klein mechanism which yet contains some of the features to be hoped for in more elaborate systems: compactification arises as a consequence of dynamics, the internal space has a non-abelian symmetry, massless chiral fermions are available, and the zero-mode sector includes no scalars. This model suffers the defects which are generally expected as well: it is not renormalizable, and the gauge couplings grow unacceptably large in the very limit which is needed to drive away the towers of massive modes. Another defect is the need for an "unnatural" adjustment of parameters in the parent lagrangian in order to flatten the four-dimensional spacetime. The presence of such adjustable parameters is both a defect and a virtue. One would like to be able to start from a less arbitrary theory, a

theory in which the "matter" content and its couplings are fixed. For this reason, the many-dimensional supergravities are attractive. Up to now, however, such theories have not yielded any physically acceptable ground state solution. Either the internal space is flat [10], with abelian symmetry or, if it is curved, it generates four-dimensional curvature of the same magnitude [5]. There are too few adjustable parameters. One hopes that this is a temporary impasse.

Finally, we believe that our approach to the phenomenon of dimensional reduction, via the computation of spectra, illuminates an important matter. This is the question of what four-dimensional fields should be included in the zero-mode ansatz, and how they should be arranged. Without first going through the spectral analysis one cannot foresee the correct form of the ansatz. Although the gauge vectors are guaranteed by the background symmetry, their placement in the ansatz has to be found. The very presence or absence of massless scalars seems to be model dependent (except when the background is supersymmetric, in which case they are certain to be present). Quite generally, the underlying mechanism whereby "matter" fields generate curvature on the internal space would seem to imply a non-trivial mixing between matter and metrical field components. This mixing has to be correctly represented in the zero-mode ansatz.

We have benefited from many stimulating and useful discussions with E. Sezgin.

Appendix

The solution to the linear eq. (3.19) which was used to obtain the pole structures (3.26)–(3.28) is listed here for reference

$$h_{ab}-\tfrac{1}{4}\eta_{ab}h_{cc}=\frac{1}{p^2+M_2^2}(T_{ab}-\tfrac{1}{4}\eta_{ab}T_{cc}),$$

$$h_{aa}=\frac{1}{(p^2+M_2^2)(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\Big[-2(p^2+M_{1-}^2)(p^2+M_{1+}^2)T_{+-}$$
$$+\frac{p^2}{a^2}T_{aa}+\frac{2}{a}M_2(p^2+M_0^2)(J_+-J_-)+\frac{2M_0M_2}{a^2}(T_{++}+T_{--})\Big],$$

$$h_{+-}=\frac{1}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\Big[\tfrac{1}{4}(p^2+M_2^2)(2T_{+-}-T_{aa})$$
$$-\frac{3}{2}\frac{M_2}{a}(J_+-J_-)-\frac{3}{2}\frac{M_0M_2}{a^2(p^2+M_0^2)}(T_{++}+T_{--})\Big],$$

$$h_{++}=\frac{1}{p^2+M_0^2}T_{++},$$

$$V_a = \frac{\sqrt{2}M_2 p_a}{a^2(p^2+M_2^2)(p^2+M_{1+}^2)(p^2+M_{1-}^2)}\left[\frac{-M_0}{a(p^2+M_0^2)}(T_{++}-T_{--})-(J_++J_-)\right]$$
$$+\frac{1}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}\left[\frac{M_2}{a}(T_{a+}-T_{a-})+(p^2+M_2^2)J_a\right],$$

$$2V_- = \frac{1}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\left[\frac{M_2}{a}(T_{+-}-\tfrac{1}{2}T_{aa})-\left(p^2+M_2^2+\frac{1}{a^2}\right)(J_+-J_-)\right.$$
$$+\frac{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}{p^2+M_2^2}(J_++J_-)-\frac{M_0}{a}\frac{p^2+M_2^2+1/a^2}{p^2+M_0^2}(T_{++}+T_{--})$$
$$\left.+\frac{M_0}{a}\frac{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}{(p^2+M_0^2)(p^2+M_2^2)}(T_{++}-T_{--})\right],$$

$$2h_{a-} = \frac{-\sqrt{2}M_2 p_a}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)}\left[T_{+-}+\frac{1}{a^2(p^2+M_2^2)}T_{bb}\right]$$
$$+\frac{1}{a(p^2+M_{1+}^2)(p^2+M_{1-}^2)}[-2M_2J_a+\sqrt{2}p_a(J_++J_-)]$$
$$+\frac{\sqrt{2}M_0 p_a}{a^2(p^2+M_{1+}^2)(p^2+M_{1-}^2)(p^2+M_0^2)}(T_{++}-T_{--})$$
$$-\frac{M_0}{\sqrt{2}}\frac{(p^2+M_2^2+1/a^2)p_a}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_0^2)}(T_{++}+T_{--})$$
$$+\left[\left(\frac{-(p^2+M_2^2)}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}+\frac{1}{p^2+M_2^2}\right)\eta_{ab}\right.$$
$$\left.-\frac{(p^2+M_2^2+1/a^2)p_a p_b}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_2^2)}\right]T_{b+}$$
$$+\left[\left(\frac{(p^2+M_2^2)}{(p^2+M_{1+}^2)(p^2+M_{1-}^2)}+\frac{1}{p^2+M_2^2}\right)\eta_{ab}\right.$$
$$\left.-\frac{(p^2+M_2^2+1/a^2)p_a p_b}{(p^2+M_{0+}^2)(p^2+M_{0-}^2)(p^2+M_2^2)}\right]T_{b-}.$$

References

[1] Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B127 (1977) 57
[2] L. Palla, Tokyo Conf. Report, (1978) p. 629
[3] S. Randjbar-Daemi and R. Percacci, Phys. Lett. 117B (1982) 41
[3a] G. Chapline and R. Slansky, Los Alamos preprint LA 82-1076
[4] A. Salam and J. Strathdee, Ann. of Phys. 141 (1982) 316

[5] P.G.O. Freund and M. Rubin, Phys. Lett. 97B (1980) 233
[6] E. Sezgin, in preparation
[7] L.F. Abbot and S. Deser, Nucl. Phys. B195 (1982) 76
[8] P. van Nieuwenhuizen and S. Deser, Phys. Rev. D10 (1974) 401
[9] N.Ya. Vilenkin, Special functions and the theory of group representations, Providence, R.I. (1968)
[10] R. Percacci and S. Randjbar-Daemi, Kaluza–Klein theories on bundles with homogeneous fibres-I, ICTP, Trieste, preprint IC/82/18, J. Math. Phys. to appear
[11] J. Scherk and J.H. Schwarz, Nucl. Phys. B153 (1979) 61

Volume 132B, number 1,2,3 PHYSICS LETTERS 24 November 1983

INSTANTON INDUCED COMPACTIFICATION AND FERMION CHIRALITY

S. RANDJBAR-DAEMI
International Centre for Theoretical Physics, Trieste, Italy
and Institut für Theoretische Physik, Boltzmangasse 5, A-1090 Vienna, Austria

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 26 August 1983

The question of fermion chirality in Kaluza–Klein theories with coupling to Yang–Mills fields is discussed. The argument is illustrated in eight dimensions where an SU(2) Yang–Mills field assumes the one-instanton form on the internal space. This serves not only to trigger spontaneous compactification of the internal space but will ensure the emergence of $n_L - n_R = \frac{2}{3}t(t+1)(2t+1)$ zero modes in an irreducible eight-spinor belonging to the $(2t+1)$-dimensional representation of SU(2).

1. One of the outstanding problems of spontaneously compactified higher dimensional field theories is the determination of the spectrum of massless fermions in the effective four-dimensional theory [1,2]. It is possible to ensure the existence of such fermionic zero-modes by coupling to gauge fields with non-trivial topology. This has been pointed out before [3] and, in a previous paper, we have applied the idea to the case of six-dimensional Einstein–Maxwell theory coupled to fermions [4]. We computed the spectra and verified the presence of fermions in complex representations of the effective local symmetry, SU(2) X U(1). The source of this chirality was the non-vanishing ground state value of the Maxwell field on the "internal" space, a two-sphere, where it assumed a magnetic monopole configuration [‡1]. The non-vanishing gauge field triggers spontaneous compactification by providing the stress components which are needed to generate the curvature of the two-sphere [5]. The non-vanishing Chern character ensures its stability [3], as well as the fermion chirality.

The aim of this note is to provide another example of the mechanism: eight-dimensional gravity coupled to SU(2) Yang–Mills fields [‡2]. Spontaneous compactification results in an internal four-sphere on which the SU(2) gauge field takes the one-instanton configuration. This configuration is SO(5)-invariant and the effective four-dimensional theory which emerges has local SO(5) symmetry.

On the issue of introducing elementary gauge fields in the higher dimensional action (which can lead, not only to stable compactification but also play a role in ensuring the chirality of the fermionic spectrum in the effective four-dimensional theory) we adopt the point of view that these could perhaps naturally arise in a supergravity theory, starting from a still higher dimension.

[‡1] The effect is possibly only if the compactified internal space has an even number of dimensions.
[‡2] This theory is free of anomalies after fermions are introduced. We are grateful to E. Witten for a discussion on anomalies in Kaluza–Klein theories.

Reprinted with permission from *Phys. Lett.* **132B** (1983) pp. 56–60
© 1983 North-Holland Publishing Company

Volume 132B, number 1,2,3 PHYSICS LETTERS 24 November 1983

2. Consider compactification induced by matter fields. Quite generally, start from the $(4+d)$-dimensional Einstein field equations coupled to matter (metric: $-1,1,1,1,...,1$) [‡3]:

$$R_{AB} = -\tfrac{1}{2}\kappa^2\{T_{AB} - [1/(2+d)]\eta_{AB}T + [2/(2+d)]\lambda\eta_{AB}\}. \tag{1}$$

Here T_{AB} is the energy–momentum tensor of the non-gravitational fields $T = \eta^{AB}T_{AB}$ and λ is a constant. We are interested in those solutions of eq. (1) which split the $(4+d)$-dimensional space into a product space $M_4 \times B_d$, where M_4 is Minkowski space–time and B_d is a compact, riemannian manifold. If we substitute $R_{ab} = 0$ in eq. (1) we get

$$T_{ab} - [1/(2+d)]\eta_{ab}T + [2/(2+d)]\lambda\eta_{ab} = 0. \tag{2}$$

From four-dimensional Lorentz invariance we conclude that

$$T_{ab} = (c/\kappa^4)\eta_{ab}, \tag{3}$$

where c is a constant. Now the assumption that the internal d-dimensional space is an Einstein space implies that

$$T_{\alpha\beta} = (c'/\kappa^4)\delta_{\alpha\beta}, \quad c' = \text{constant}. \tag{4}$$

Upon substitution of (3) and (4) into (2) and into (α,β) components of (1), we get

$$\lambda\kappa^4 = c + (c'-c)\tfrac{1}{2}d, \quad R_{\alpha\beta} = (-1/2\kappa^2)(c'-c)\delta_{\alpha\beta}. \tag{5,6}$$

For a symmetric riemannian space, compactness is ensured if $c' - c > 0$. These constants will depend on the details of the individual models [‡4]. For the case of gauge fields alone, due to Lorentz invariance, eqs. (3)–(6) reduce to

$$\lambda = \tfrac{1}{4}\bar{F}^2, \quad R_{\alpha\beta} = (-\kappa^2\bar{F}^2/2d)\delta_{\alpha\beta}, \quad c = -\tfrac{1}{4}\kappa^4 F^2, \quad c' = \kappa^4\bar{F}^2(4-d)/4d, \tag{7}$$

where $F^2 = F_{\alpha\beta}\cdot F^{\alpha\beta}$. The negativity of the Ricci tensor, $R_{\alpha\beta}$, ensures compactification. Thus the problem reduces to finding appropriate solutions of the Yang–Mills equations in the internal space. These solutions must have a constant F^2. In the case of a homogeneous internal space, G/H, there exist standard solutions of the gauge field equations, where all components of the Yang–Mills fields are zero, except those which belong to H or any of its subgroups. (Recall that H must be contained in the Yang–Mills gauge group.) Such solutions have been discussed in ref. [7]; they have the characteristic property that $T_{\alpha\beta}$ is proportional to $(d-4)$ and vanishes when the internal space is four-dimensional (as for example in the case of the internal BPST instanton).

3. Now consider the case of eight-dimensional gravity coupled to SU(2) Yang–Mills. One can demonstrate the existence of a vacuum solution with the geometry $M^4 \times S^4$, where the four-dimensional internal space, S^4, supports an SU(2) instanton. We would like to establish the local symmetry of this background by identifying the massless vector states.

It is well known that the single instanton solution on S^4 is invariant under the combined action of O(5) and SU(2) [8]. More precisely, under the isometric action of O(5) on S^4 the instanton configuration changes by an SU(2) gauge transformation. Therefore it is natural to classify the spectrum of small oscillations into multiplets of O(5). To start with the spectral analysis we substitute

$$g_{MN} = \bar{g}_{MN} + h_{MN}, \quad A_M = \bar{A}_M + V_M, \tag{8}$$

into the action integral and expand it in a power series of h_{MN} and V_M. Here $\bar{g}_{MN}$ and $\bar{A}_M$ indicate the background classical solution and h_{MN} and V_M are small perturbations which depend on all the eight coordinates in a com-

‡3 Our notation is the same as in ref. [4]. Unless otherwise stated we use an orthonormal basis in the manifold.

‡4 The source of compactification may be a quantum loop in which case one would have to make a self-consistent computation [6] of these.

pletely arbitrary way. As a result of the general coordinate and the SU(2)-gauge invariance of the eight-dimensional theory, the linearized theory will be invariant under the following set of local gauge transformations:

$$\delta h_{MN} = \xi_{M\cdot N} + \xi_{N\cdot M}, \quad \delta V_M = \Omega_{\cdot M} + F_{MN}\xi^N, \tag{9}$$

where ξ_M and Ω are arbitrary parameters and "$\cdot N$" refers to the covariant differentiation relative to the background connection. To make the theory well defined we impose the following gauge conditions:

$$(h_{MN} - \tfrac{1}{2}\bar{g}_{MN}h_{PP})_{\cdot M} = 0, \quad V_{N\cdot M} = 0. \tag{10}$$

Subject to these constraints, the spin-one modes in the bilinear action are given by

$$\begin{aligned} S_2(\text{spin-one}) = \frac{V(S^4)}{\kappa^4}\int d^4x \Big(\sum_{n\geq 1}\sum_q \{ & \tfrac{1}{2}h^{n1}_{a,q}(x)[\partial^2 - (n/a^2)(n+3) - 2/a^2]h^{n1}_{a,q}(x) \\ & + \tfrac{1}{2}h^{n0}_{a,q}(x)[\partial^2 - (n/a^2)(n+3)]h^{n0}_{a,q}(x) + \tfrac{1}{2}V^{n1}_{a,q}(x)[\partial^2 - (n/a^2)(n+3) + 2/a^2]V^{n1}_{a,q}(x) \\ & - (\sqrt{2}/a^2)[(n+1)(n+2)]^{1/2}h^{n1}_a V^{n1}_a\} \Big). \end{aligned} \tag{11}$$

Here (n_1, n_2) specify the O(5)-representation of dimension $d(n_1, n_2)$ given by

$$d(n_1, n_2) = \tfrac{1}{6}(n_1 + n_2 + 2)(n_1 - n_2 + 1)(2n_1 + 3)(2n_2 + 1), \tag{12}$$

and q can be assumed to range from 1 to $d(n_1, n_2)$.

It is easily seen that S_2 (spin-one) contains two classes of representations characterized by $(n, 0)$ and $(n, 1)$, with $n \geq 1$. In the first class all modes are massive with their (mass)^2 given by

$$(n/a^2)(n+3) \quad (n \geq 1), \tag{13}$$

while in the second class the (mass)^2 matrix is as follows:

$$M_n^2 = \begin{pmatrix} (n/a^2)(n+3) + 2/a^2 & (\sqrt{2}/a^2)[(n+1)(n+2)]^{1/2} \\ (\sqrt{2}/a^2)[(n+1)(n+2)]^{1/2} & (n/a^2)(n+3) - 2/a^2 \end{pmatrix}. \tag{14}$$

It follows that

$$\det M_n^2 = a^{-4}(n-1)(n^3 + 7n^2 + 14n + 6). \tag{15}$$

Since $n \geq 1$, the *only* zero of $\det M_n^2$ is located at $n = 1$. The formula for $d(n_1, n_2)$ indicates that the multiplicity of this mode is ten. The massless spin-one fields are thereby seen to belong to the adjoint representation of O(5). We conclude that the effective four-dimensional theory is an O(5) Yang–Mills theory. Notice that the SU(2) Yang–Mills symmetry of the eight-dimensional lagrangian is spontaneously broken [‡5]. From (11) it is clear that the spin-one sector is stable against small oscillations.

4. We now consider the question of chirality. Classical backgrounds can be topologically non-trivial and may therefore induce chiral fermion spectra. More precisely, it can be asserted that, if the internal space is even-dimensional then there will emerge a set of fermionic zero-modes whose number is governed by the Atiyah–Singer index theorem [9]. To see this consider the Dirac lagrangian in $(4 + d)$-dimensional space–time,

$$\mathcal{L}_D = i\bar{\psi}\not{\nabla}\psi, \tag{16}$$

[‡5] This point was overlooked in ref. [7] where the statement with regard to the number of massless spin-one fields is incorrect.

Volume 132B, number 1,2,3 PHYSICS LETTERS 24 November 1983

in which $\not{\nabla} = \Gamma^A \nabla_A = \Gamma^A (e_A + \omega_A + A_A)$ where ω_A and A_A denote the Riemann and Yang–Mills connections, respectively.

We shall assume that d is *even* and consider the following realization of the Dirac matrices, Γ_A,

$$\Gamma_a = \gamma_a \times 1, \quad a = 0,1,2,3; \quad \Gamma_\alpha = \gamma_5 \times \gamma'_\alpha, \quad \alpha = 5,6,...,4+d, \tag{17}$$

where γ_a and γ'_α are the four- and d-dimensional Dirac matrices, respectively. This realization is appropriate for treating the decomposition of the SO(1,3 + d) spinor under the subgroup SO(1,3) × SO(d), the product of the tangent space groups on the Minkowski and internal spaces. The two eigenvalues of the SO(1,3 + d) invariant matrix,

$$\bar{\Gamma} = \Gamma_0 \Gamma_1 \ldots \Gamma_{4+d}, \tag{18}$$

distinguish two inequivalent irreducible representations of the tangent space group, SO(1,3 + d). In the representation (17) we have

$$\bar{\Gamma} = \gamma_5 \times \bar{\gamma}', \tag{19}$$

where $\bar{\gamma}' = \gamma'_5 \gamma'_6 \ldots \gamma'_{4+d}$ is SO(d) invariant, and γ_5 is the familiar SO(1,3) invariant which distinguishes chirality. It is clear that an irreducible spinor of SO(1,3 + d), say $\bar{\Gamma} = 1$, must branch into two irreducible pieces under the subgroup SO(1,3) × SO(d),

$$\psi_L \quad \text{with } \gamma_5 = 1 \quad \text{and } \bar{\gamma}' = 1, \quad \psi_R \quad \text{with } \gamma_5 = -1 \quad \text{and } \bar{\gamma}' = -1. \tag{20}$$

The main point to emphasize is that the left and right components belong to *inequivalent* representations of SO(d). (It can be shown [‡6] that the representations corresponding to $\bar{\gamma}' = 1$ and $\bar{\gamma}' = -1$ are conjugate complex if $d = 4k + 2$, and real or pseudoreal but inequivalent if $d = 4k$.) The theory possesses a *global* U(1) invariance.

The mass operator takes the form

$$\bar{\psi}_L \, i\Gamma^\alpha \nabla_\alpha \psi_R + \text{h.c.}, \tag{21}$$

where ∇_α operates in the internal space. This operator will have zero eigenvalues as the index theorem dictates. For example, in the model discussed above where the internal space is the four-dimensional sphere, S^4, and the background gauge field takes an instanton configuration, let ψ belong to the $(2t + 1)$-dimensional representation of the Yang–Mills SU(2). Then

$$N_L - N_R = \tfrac{2}{3} t(t+1)(2t+1)k, \tag{22}$$

where N_L and N_R are the numbers of left- and right-handed zero modes and k is the instanton number.

The modes can be found explicitly if the internal space is a quotient space G/H and the gauge field configuration is G-invariant. The modes belong to irreducible representations of G and the zero modes correspond to those representations of G which appear in the harmonic expansion of either ψ_L or ψ_R (but not both).

In the above example, if $k = 1$ then the background has SO(5) invariance and the zero modes must belong to a representation of this group. To find the zero-modes it is necessary only to examine the harmonic expansions of spinors on S^4. A general technique for constructing these expansions has been described elsewhere [11]. It involves decomposing the spinor which belongs to an irreducible representation of the group $SO(1,7)_{\text{tangent space}} \times SU(2)_{\text{gauge}}$, into pieces according to the chain

$$SO(1,7)_{ts} \times SU(2)_g \to (SL(2,\mathbf{C}) \times SU(2)_A \times SU(2)_B)_{ts} \times SU(2)_g \to SL(2,\mathbf{C}) \times SU(2)_A \times SU(2)_{B+g}.$$

That is, from the tangent space group (or rather its covering group) is picked out the subgroup associated with the product manifold $M^4 \times S^4$. The factor associated with S^4 has been denoted as $SU(2)_A \times SU(2)_B$. the instantion is supposed to lie in the algebra of $SU(2)_B$. It is invariant with respect to the diagonal subgroup of $SU(2)_B$

‡6 See for example the discussion in ref. [9].

$\times$ SU(2)$_g$ in which tangent space rotations are accompanied by gauge transformations. Finally, the pieces which are irreducible with respect to SU(2)$_A$ $\times$ SU(2)$_{B+g}$ can be expanded in SO(5) harmonics.

Now consider the eight-spinor of SO(1,7) which also belongs to $2t+1$ of SU(2)$_g$. With respect to SL(2, C) $\times$ SU(2)$_A$ $\times$ SU(2)$_B$ it branches into two pieces, ψ_L and ψ_R. The SU(2)$_A$ $\times$ SU(2)$_B$ $\times$ SU(2)$_g$ content is

$$\psi_L \sim (2,1,2t+1) \quad \text{and} \quad \psi_R \sim (1,2,2t+1). \tag{23}$$

With respect to SU(2)$_A$ $\times$ SU(2)$_{B+g}$, we have

$$\psi_L \sim (2,2t+1), \quad \psi_R \sim (1,2t)+(1,2t+2). \tag{24}$$

To facilitate the expansion we re-express these SO(4) representations in the notation $[m_1,m_2]$ of Gel'fand and Zetlin [12],

$$\psi_L \sim [t+1/2,t-1/2], \quad \psi_R \sim [t-1/2,t-1/2]+[t+1/2,t+1/2]. \tag{25}$$

The SO(5)-harmonics are labelled by a pair of integers (or half integers) (n_1,n_2) and the expansion of an SO(4) piece $[m_1,m_2]$ must include all (n_1,n_2) subject to the inequalities

$$n_1 \geqslant m_1 \geqslant n_2 \geqslant |m_2|. \tag{26}$$

It is easy to verify that the only harmonic which fails to appear in both ψ_L and ψ_R is

$$n_1 = n_2 = t - 1/2, \tag{27}$$

which is present only in ψ_R. This must constitute the zero-mode multiplet. Its dimensionality is given by eq. (20)

$$d(t-1/2,t-1/2) = \tfrac{2}{3}t(t+1)(2t+1), \tag{28}$$

which agrees with the prediction (22) of the index theorem.

We are appreciative of discussions with E. Witten.

References

[1] A. Salam, Proc. 19th Intern. Conf. on High energy physics (Tokyo, 1978) p. 597.
[2] E. Witten, Nucl. Phys. B186 (1981) 412.
[3] Z. Horvath, L. Palla, E. Cremmer and J. Scherk, Nucl. Phys. B127 (1977) 57;
L. Palla, Proc. 19th Intern. Conf. on High energy physics (Tokyo, 1978) p. 629.
[4] S. Randjbar-Daemi, A. Salam and J. Strathdee, Nucl. Phys. B214 (1983) 491.
[5] E. Cremmer and J. Scherk, Nucl. Phys. B108 (1976) 408;
J.F. Luciani, Nucl. Phys. B135 (1978) 111.
[6] P. Candelas and S. Weinberg, Texas preprint UTTG-6-83.
[7] S. Randjbar-Daemi and R. Percacci, Phys. Lett. 117B (1982) 41.
[8] R. Jackiw and C. Rebbi, Phys. Rev. D16 (1977) 1052.
[9] T. Eguchi, P.B. Gilkey and A.J. Hanson, Phys. Rep. 66 (1980) 213.
[10] G. Chapline and R. Slansky, Nucl. Phys. B209 (1982) 461.
[11] A. Salam and J. Strathdee, Ann. Phys. 141 (1982) 316.
[12] A.O. Barut and R. Raczka, Theory of group representations and applications (Polish Scientific Publishers, Warsaw, 1980).

Volume 151B, number 5,6 PHYSICS LETTERS 21 February 1985

AN ANOMALY-FREE MODEL IN SIX DIMENSIONS

S. RANDJBAR-DAEMI [1]
Institute for Theoretical Physics, Sidlerstrasse 5, CH-3012 Bern, Switzerland

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

and

E. SEZGIN and J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 28 November 1984

We show that a gauged supergravity theory based on $E_6 \times E_7 \times U(1)$ is free of gauge and gravitational anomalies in six dimensions. It compactifies to $(\text{Minkowski})^4 \times S^2$ by the standard monopole mechanism. With a monopole of strength n in E_6, the resulting four-dimensional theory exhibits chiral $SO(10) \times U(1)$ with $2|n|$ families (and no antifamilies). Supersymmetry is broken.

1. Introduction. Green and Schwarz [1] have recently shown that ten-dimensional $N = 1$ supergravity coupled to O(32) or $E_8 \times E_8$ Yang–Mills fields is free of gauge and gravitational anomalies. Witten [2] has argued that the O(32) model might even be realistic in that it could compactify so as to yield a four-dimensional effective chiral SU(5) gauge theory with any number of generations of quarks and leptons. Such a compactification however is conjectural [3]. If solutions can be found it would be interesting to know whether they correspond to minkowskian or anti-de Sitter space–time, and whether they are stable.

In this note we give criteria for constructing anomaly free gauge supergravities in six dimensions. These theories involve an antisymmetric tensor as part of the gravity supermultiplet and this field plays the same role in the structure of anomaly cancelling counterterms as in the ten-dimensional model of Green and Schwarz. In six dimensions it is possible to introduce matter as well as gauge multiplets and, as we show, both are necessary in order to achieve cancellation of gauge and gravitational anomalies. We also exhibit explicit compactification on Minkowski $\times S^2$.

An important property of O(32), exploited by Green and Schwarz is the absence of an algebraically independent sixth-order Casimir invariant [4], i.e. $\operatorname{Tr} F^6 \sim \operatorname{Tr} F^2 \operatorname{Tr} F^4$. For six dimensions the analogous property, which we use, is the absence of independent fourth-order invariants in the exceptional algebras [3] [‡1]. Apart from this, the structure of our model is heavily influenced by the purely gravitational anomalies. Specifically, the various contributions to the term $\operatorname{tr} R^4$ are required to cancel.

One of the aims in constructing the model was to find an explicit compactifying solution with the geometry of Minkowski space–time ⊗ internal two-sphere. One way to achieve such compactification is through a non-vanishing magnetic monopole background configuration on the two-sphere. It is therefore required

[1] Supported by Schweizerischer Nationalfonds.

[‡1] Note that SU(2) and SU(3) have the same property. A model based on $SU(3) \times SU(2) \times U(1)$ is being considered.

Reprinted with permission from *Phys. Lett.* **151B** (1985) pp. 351–356
© 1985 Elsevier Science Publishers B.V.

that the model contain gauge fields. Among various anomaly free possibilities we found that the most elegant combination involved $E_6 \times E_7 \times U(1)$. The monopole background lies in the algebra of E_6 causing it to break spontaneously to SO(10) in the course of compactification. This compactification leads to a chiral SO(10) theory in four dimensions with the number of quark and lepton generations, $2|n|$, governed by the monopole strength, n. (Stability at the classical level requires $|n| = 1$.) Supersymmetry is broken by the E_6 background. The gravitino, along with the hypermatter is massive. The E_7 gauge symmetry remains unbroken but all light fermions are neutral with respect to it.

2. The model. In six dimensions the following $N = 2$ supermultiplets exist [5]:

gravity $e_\mu{}^m, \psi_{\mu L}^A, B_{\mu\nu}^+, \quad \mu, m = 0, 1, ..., 5, \ A = 1, 2,$

hypermatter $\psi_R{}^a, \phi^\alpha, \quad a = 1, ..., 2n, \ \alpha = 1, ..., 4n,$

Yang–Mills $A_\mu, \lambda_L{}^A,$

tensor $B_{\mu\nu}^-, \chi_R{}^A, \sigma .$

The antisymmetric tensors $B_{\mu\nu}^+$ and $B_{\mu\nu}^-$ are constrained to have self-dual and anti-self-dual field strengths, respectively. In the model to be treated here they appear only through the combination $B_{\mu\nu}^+ + B_{\mu\nu}^-$, which is unconstrained. The index $A = 1, 2$ indicates that $\psi_{\mu L}$, λ_L and χ_R are doublets with respect to Sp(1). They are also Weyl spinors ψ_L or ψ_R with respect to SO(1, 5). They belong to *real* representations of SO(1, 5) × Sp(1) and are subject to reality conditions. In our model the Sp(1) symmetry is explicitly broken to U(1), which is gauged. The hyperinos ψ_R^a are neutral with respect to U(1) although their partners ϕ^α are charged. In general, the hyperscalars ϕ^α – with $2n$ complex or $4n$ real components – must parametrize a non-compact quarternionic Kähler manifold [6].

In our model this turns out to be Sp(456, 1)/Sp(456) × Sp(1). The hypermatter belongs to the pseudoreal **912** of Sp(456) and the fermions are subject to a reality condition of the symplectic-Majorana type [‡2]. The dimensionality, $4n = 4 \times 456$, of the hyper-Kähler manifold is determined here by the requirement that the gravitational anomaly of the type $\mathrm{tr}\, R^4$ should be absent. It can be shown that this happens only if the number, n, of hyperinos is related to the number, m, of gauginos by [‡3]

$$n = m + 244 . \tag{2.1}$$

We take $m = 78 + 133 + 1$ corresponding to the Yang–Mills group $E_6 \times E_7 \times U(1)$. With $n = 456$ it is then natural to assign the hyperinos to the 912 of $E_7 \subset$ Sp(456). What is then truly remarkable is the complete cancellation of all gauge and gravitational anomalies which results (see section 3).

Locally supersymmetric couplings of the above multiplets have been given in ref. [5]. In particular, the antisymmetric tensor was included in order to overcome the difficulties associated with the action principle for constrained antisymmetric tensors. In ref. [5], the entire hypersymmetry Sp(n) was gauged. Here we shall gauge only the subgroup E_7 with respect to which the 912 of Sp(456) is irreducible.

The lagrangian is given by [‡4]

$$\begin{aligned}(-g)^{-1/2}\mathcal{L} = {} & \tfrac{1}{4}\kappa^{-2}R - \tfrac{1}{4}(\partial_\mu\sigma)^2 - \tfrac{1}{12}\kappa^2 e^{2\kappa\sigma} G_{\mu\nu\rho}^2 \\ & - \tfrac{1}{4} e^{\kappa\sigma}(g_6^{-2} F_{6\mu\nu}^{I\,2} + g_7^{-2} F_{7\mu\nu}^{I2} + g_1^{-2} F_{1\mu\nu}^2) \\ & - \tfrac{1}{2} g_{\alpha\beta}(\phi)\, D_\mu\phi^\alpha\, D_\mu\phi^\beta \\ & - \tfrac{1}{8}\kappa^{-4} e^{-\kappa\sigma}(g_7^2 C_{7i}^{I2} + g_1^2 C_{1i}^2) + \text{fermion terms} ,\end{aligned} \tag{2.2}$$

where $g_{\alpha\beta}(\phi)$ is the quarternionic-Kähler metric, and the usual riemannian $g_{\mu\nu}(x)$ is tacit. The gauge field strengths and coupling constants of $E_6 \times E_7 \times U(1)$ are denoted, respectively, by F_6, F_7, F_1 and g_6, g_7, g_1. The hyperscalar potential is constructed from

$$\begin{aligned} C_{7i}^I &= A_\alpha^i(\phi)(T^I\phi)^\alpha , \quad i = 1, 2, 3 , \\ C_{1i} &= A_\alpha^i(\phi)(T^3\phi)^\alpha - \delta_{i3} , \end{aligned} \tag{2.3}$$

[‡2] Other non-compact quarternionic Kähler manifolds are the symmetric spaces $SU(n, 2)/SU(n) \times U(1) \times Sp(1)$, $SO(n, 4)/SO(n) \times SO(3) \times Sp(1)$, $G_2/SO(3) \times Sp(1)$, $F_4/Sp(3) \times Sp(1)$, $E_6/SU(6) \times Sp(1)$, $E_7/SO(12) \times Sp(1)$, $E_8/E_7 \times Sp(1)$. The associated hyperinos would be assigned, respectively, to the real representation $(n, 1)_1 + (\bar{n}, 1)_{-1}$, $(n, 2, 1)$, $(4, 1)$, $(14, 1)$, $(20, 1)$, $(32, 1)$ and $(56, 1)$ of the appropriate isotropy group. A model based on $E_7/SO(12) \times Sp(1)$ has been considered in ref. [7], but is anomalous.

[‡3] This condition (2.1) changes to $n = m + 273 - 29k$, if k, the number of antisymmetric $B_{\mu\nu}^-$ fields, does not equal unity. Theories with $k \neq 1$ would not have a covariant lagrangian formulator, only covariant field equations. In this note we concentrate on $k = 1$.

[‡4] See appendix for fermion terms.

where $A^i_\alpha(\phi)$ denotes the Sp(1) part of the canonical connection on the hyper-Kähler manifold. The generators of $E_7 \times Sp(1)$ are denoted by T^I and $T^i = -i\sigma^i/2$. The field strength tensor $G_{\mu\nu\rho}$ includes the curl of the antisymmetric tensor $B_{\mu\nu}$ and the Chern–Simons forms of the gauge group, $E_6 \times E_7 \times U(1)$. In the notation of differential forms,

$$G = dB + g^{-2}\,\mathrm{tr}(A\,dA - \tfrac{2}{3}A^3)\,. \tag{2.4}$$

The presence of the non-negative hyperscalar potential in (2.2) is important for the minkowskian compactification of the theory [8].

3. Anomaly cancellation. In six-dimensional space–time the various anomalies are associated with quadrangle graphs rather than triangles as in four dimensions. Another important difference is that the Weyl spinors of SO(1, 5) are pseudoreal rather than complex. If anomalies are to be cancelled then fields of both chiral types must be involved. In the model discussed here the fermions and their classification with respect to $SO(1,5) \times E_6 \times E_7 \times U(1)$ are as follows: gravitino $\psi_{\mu L} \sim (4_+, 1, 1)_1$; hyperino $\psi_R \sim (4_-, 1, 912)_0$; gaugino $\lambda_L \sim (4_+, 78, 1)_1 + (4_+, 1, 133)_1 + (4_+, 1, 1)_1$; tensorino $\chi_R \sim (4_-, 1, 1)_1$, where the suffixes L and R correspond to the Weyl spinors 4_+ and 4_-, respectively. Because of the U(1) charge (indicated by a suffix) all these fields are complex except for ψ_R which is real (because the representation 912 of E_7 is pseudoreal).

The quadrangle anomalies are usefully codified in the form of an invariant polynomial [9], $P(F, R)$, which is an eight-form made from the Yang–Mills and curvature two-forms. The coefficients in this polynomial receive contributions from the fermions (listed above). According to the work of Alvarez-Gaumé and Witten [10] the various contributions are given by

$$\begin{aligned}
P(\psi_\mu) &= \tfrac{5}{24}F_1^4 - \tfrac{19}{96}F_1^2\,\mathrm{tr}\,R^2 + \tfrac{245}{5760}\,\mathrm{tr}\,R^4 \\
&\quad - [(5\times 43)/(4\times 5760)](\mathrm{tr}\,R^2)^2\,, \\
-2P(\psi_R) &= \tfrac{1}{24}\mathrm{Tr}_{912}F_7^4 + \tfrac{1}{96}\mathrm{Tr}_{912}F_7^2\,\mathrm{tr}\,R^2 \\
&\quad + \tfrac{912}{5760}[\mathrm{tr}\,R^4 - \tfrac{5}{4}(\mathrm{tr}\,R^2)^2]\,, \\
-P(\chi_R) &= \tfrac{1}{24}F_1^4 + \tfrac{1}{96}F_1^2\,\mathrm{Tr}\,R^2 \\
&\quad + \tfrac{1}{5760}[\mathrm{tr}\,R^4 - \tfrac{5}{4}(\mathrm{tr}\,R^2)^2]\,,
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
P(\lambda_L) &= \tfrac{1}{24}(\mathrm{Tr}_{78}\,F_6^4 + 6\,\mathrm{Tr}_{78}F_6^2F_1^2 + 78\,F_1^4) \\
&\quad + \tfrac{1}{24}(\mathrm{Tr}_{133}F_7^4 + 6\,\mathrm{Tr}_{133}F_7^2F_1^2 + 133\,F_1^4) + \tfrac{1}{24}F_1^4 \\
&\quad + \tfrac{1}{96}[\mathrm{Tr}_{78}F_6^2 + \mathrm{Tr}_{133}F_7^2 + (78+133+1)F_1^2]\,\mathrm{tr}\,R^2 \\
&\quad + [(78+133+1)/5760][\mathrm{tr}\,R^4 - \tfrac{5}{4}(\mathrm{tr}\,R^2)^2]\,,
\end{aligned} \tag{3.1 cont'd}$$

where the field strength two-forms corresponding to O(1, 5), E_6, E_7 and U(1) are denoted by R, F_6, F_7 and F_1, respectively. The traces involving R are in the six-dimensional representation of O(1, 5) while the other traces are specified by a subscript to indicate the representation. In the case of ψ_R an extra factor of $\frac{1}{2}$ is needed to take account of the reality condition on ψ_R.

Because E_6 and E_7 have no independent quartic invariant it is possible to make some simplifications [3]:

$$\begin{aligned}
&\mathrm{Tr}_{78}F_6^4 = \tfrac{1}{2}(\mathrm{Tr}_{27}F_6^2)^2, \quad \mathrm{Tr}_{78}F_6^2 = 4\,\mathrm{Tr}_{27}F_6^2\,, \\
&\mathrm{Tr}_{133}F_7^4 = \tfrac{1}{6}(\mathrm{Tr}_{56}F_7^2)^2, \quad \mathrm{Tr}_{133}F_7^2 = 3\,\mathrm{Tr}_{56}F_7^2, \\
&\mathrm{Tr}_{912}F_7^2 = \tfrac{31}{12}(\mathrm{Tr}_{56}F_7^2)^2, \quad \mathrm{Tr}_{912}F_7^2 = 30\,\mathrm{Tr}_{56}F_7^2\,.
\end{aligned}$$

The total P then reduces to

$$\begin{aligned}
P &= -\tfrac{1}{16}(\mathrm{tr}\,R^2)^2 + \tfrac{1}{24}\mathrm{tr}\,R^2\,\mathrm{Tr}_{27}F_6^2 \\
&\quad - \tfrac{1}{16}\mathrm{tr}\,R^2\,\mathrm{Tr}_{56}F_7^2 + 2\,\mathrm{tr}\,R^2\,F_1^2 + \tfrac{1}{48}(\mathrm{Tr}_{27}F_6^2)^2 \\
&\quad + \mathrm{Tr}_{27}F_6^2F_1^2 - \tfrac{3}{64}(\mathrm{Tr}_{56}F_7^2)^2 \\
&\quad + \tfrac{3}{4}\,\mathrm{Tr}_{56}F_7^2F_1^2 + 9F_1^4\,.
\end{aligned} \tag{3.2}$$

The $\mathrm{tr}\,R^4$ contributions have cancelled. The four factors O(1, 5), E_6, E_7 and U(1) are involved in a rather symmetrical fashion and it proves advantageous to express P in the form

$$P = \sum_{j,k}\beta_{jk}G_jG_k\,, \tag{3.3}$$

where

$$G_1 = \tfrac{1}{4}\mathrm{tr}\,R^2, \quad G_2 = \tfrac{1}{4}\mathrm{Tr}_{27}F_6^2,$$
$$G_3 = \tfrac{3}{8}\mathrm{Tr}_{56}F_7^2, \quad G_4 = 3\,F_1^2, \tag{3.4}$$

and

$$\beta_{jk} = \tfrac{1}{3}\begin{pmatrix} -3 & 1 & -2 & 4 \\ 1 & 1 & 0 & 2 \\ -2 & 0 & -1 & 1 \\ 4 & 2 & 1 & 3 \end{pmatrix}. \tag{3.5}$$

The four-forms G_k are closed; they can be expressed locally in terms of Chern–Simons forms $G_k = \mathrm{d}\omega_k^0$. Since G_k is invariant, it follows that the gauge variation $\delta\omega_k^0$ is itself closed, i.e. $\delta\omega_k^0 = \mathrm{d}\omega_k^1$. It can be shown [9] that six-forms $G_j\,\omega_k^1$ satisfy the Wess–Zumino consistency conditions and hence that the potential anomalies corresponding to (3.3) would take the form

$$\omega_6^1 = \sum_{jk} \beta_{jk}\, G_j\, \omega_k^1 . \tag{3.6}$$

However, the anomaly is defined only up to terms of the form $\delta(\Delta\mathcal{L})$, the gauge variation of a counterterm. Our aim now is to show that the expression (3.6) can be completely eliminated by a counterterm.

What counterterms are available? Following the approach of Green and Schwarz we shall consider those of the form

$$\Delta'\mathcal{L} = \sum_j \gamma_j\, G_j\, B + \tfrac{1}{2}\sum_{jk} \epsilon_{jk}\, \omega_j^0\, \omega_k^0 , \tag{3.7}$$

where B is the two-form specific to this type of supergravity. The parameters γ_j and $\epsilon_{jk} = -\epsilon_{jk}$ are at our disposal. We shall suppose that B transforms according to

$$\delta B = -\sum_j \alpha_j\, \omega_j^1 , \tag{3.8}$$

where, again, the parameters α_j are free. The gauge variation of (3.7) is

$$\delta(\Delta'\mathcal{L}) = -\sum_{jk} (\gamma_j\, \alpha_k - \epsilon_{jk})\, G_j\, \omega_k^1 . \tag{3.9}$$

If this is to cancel the expression (3.6) we must have $\beta_{jk} = \gamma_j\,\alpha_k - \epsilon_{jk}$. But β is symmetric while ϵ is antisymmetric. Therefore

$$\epsilon_{jk} = \tfrac{1}{2}(\gamma_j\alpha_k - \gamma_k\alpha_j), \quad \beta_{jk} = \tfrac{1}{2}(\gamma_j\alpha_k + \gamma_k\alpha_j). \tag{3.10}$$

In other words, the model is free of gauge and gravitational anomalies if the matrix β given by (3.5) can be expressed in the form (3.10) with real parameters γ and α. This highly non-trivial requirement is indeed satisfied. One easily finds

$$\alpha_j = (3, 1, 1, 1), \quad \gamma_j = (-\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{1}{3}, 1). \tag{3.11}$$

The complete counterterm is given by

$$\begin{aligned} \Delta\mathcal{L} &= \sum_j \gamma_j\, G_j\, B + \tfrac{1}{2}\sum_{jk} \gamma_j\, \omega_j^0 \alpha_k\, \omega_k^0 \\ &\quad + \mathcal{L}(G + \Delta G) - \mathcal{L}(G) \\ &= (-\tfrac{1}{12}\,\mathrm{tr}\, R^2 + \tfrac{1}{12}\,\mathrm{tr}\, F_6^2 - \tfrac{1}{8}\,\mathrm{tr}\, F_7^2 + 3F_1^2)B \\ &\quad + \tfrac{1}{2}(-\tfrac{1}{3}\omega_1^0 + \tfrac{1}{3}\omega_2^0 - \tfrac{1}{3}\omega_3^0 + \omega_4^0) \\ &\quad \times (3\omega_1^0 + \omega_2^0 + \omega_3^0 + \omega_4^0) + \mathcal{L}(G + \Delta G) - \mathcal{L}(G) . \end{aligned} \tag{3.12}$$

where $\Delta G = 3\omega_1^0 - 3\omega_2^0 - \tfrac{5}{3}\omega_3^0 + \tfrac{2}{3}\omega_4^0$ and the G-containing part of $\mathcal{L}$ is

$$\begin{aligned} \mathcal{L}(G) &= -\tfrac{1}{12}\kappa^2 \mathrm{e}^{2\kappa\sigma} G_{\mu\nu\rho}^2 \\ &\quad + \tfrac{1}{24}\kappa^2 \mathrm{e}^{\kappa\sigma} G_{\mu\nu\rho}(\bar\psi^\lambda \gamma_{[\lambda}\gamma^{\mu\nu\rho}\gamma_{\tau]}\psi^\tau + 2\bar\psi^\lambda\gamma^{\mu\nu\rho}\gamma_\lambda\chi \\ &\quad - \bar\chi\gamma^{\mu\nu\rho}\chi + \bar{\hat\lambda}\gamma^{\mu\nu\rho}\hat\lambda + \bar\psi^a\gamma^{\mu\nu\rho}\psi_a) , \end{aligned} \tag{3.13}$$

where $\hat I$ labels the adjoint representation of $E_6 \times E_7 \times U(1)$.

4. Compactification. At the classical level the equations of motion admit a vacuum solution in which the six-dimensional geometry factorizes into the product of four-dimensional minkowskian space–time with an internal two-sphere. This compactification is driven by one of the gauge fields which assumes a monopole configuration on the two-sphere. All other fields either vanish or assume constant values. There are many ways to pick out the non-vanishing gauge field since we have a 212-dimensional local symmetry $E_6 \times E_7 \times U(1)$. For the purposes of this note we shall embed the monopole in the algebra of E_6, leaving $E_7 \times U(1)$ unbroken. The E_6 symmetry will itself break to SO(10). This means that the final vacuum symmetry in four dimensions will be Poincaré $\times$ $SU(2)_{\text{Kaluza–Klein}}$ $\times$ $SO(10) \times E_7 \times U(1)$. One of our tasks in the following is to isolate the zero mode sector and show that it is free of anomalies (in the usual four-dimensional sense).

First consider the relevant equations of motion derived from the bosonic part of (2.2). Since the only non-vanishing vacuum fields are the scalar σ, a U(1)

gauge field from E_6 and the metric tensor, all equations are trivially satisfied except for

$$\kappa^{-2}R_{AB} = 2g_6^{-2}e^{\kappa\sigma}(F^{I}_{AC}F^{IC}_{B}) + D_A\sigma\, D_B\sigma - \tfrac{1}{2}\kappa^{-1}g_{AB}\,\Box\sigma\,, \quad (4.1)$$

$$\kappa^{-1}\Box\sigma = \tfrac{1}{2}g_6^{-2}e^{\kappa\sigma}(F^{I}_{AB}F_{I}^{AB}) - g_1^2\tfrac{1}{4}\kappa^{-4}e^{-\kappa\sigma}, \quad (4.2)$$

$$D_A(\sqrt{-g}\,e^{\kappa\sigma}F^{AB}) = 0\,, \quad (4.3)$$

where g_6 and g_1 denote the E_6 and U(1) coupling constants, respectively, and $16\pi\kappa^2$ is the four-dimensional gravitation coupling.

We look for a solution of the form $M^4 \times S^2$ with constant σ and with the E_6 Yang–Mills field representing a monopole. The potential one-form can then be written as

$$A_\pm = \nu Q(\cos\theta \mp 1)\,d\varphi\,, \quad (4.4)$$

where θ and φ are the usual polar coordinates on the internal S^2, Q is a generator of E_6 and ν is a number. It is necessary to use two coordinate patches and two expressions for the potential, one for the northern hemisphere and one for the southern. On the equator these potentials must be connected by a single-valued gauge transformation, $A_- = S^{-1}A_+S + iS^{-1}\,dS$. A suitable mapping is $S(\varphi) = \exp(2i\nu Q\varphi)$. The important condition of single valuedness, $S(2\pi) = S(0)$, implies that $2\nu Q$ must be an integer.

We now identify Q with the U'(1) factor in the subgroup $SO(10) \times U'(1)$ of E_6. It can be normalized such that the 27 of E_6 branches according to

$$27 = 16_{1/3} + 10_{-2/3} + 1_{4/3}\,,$$

while the adjoint representation gives

$$78 = 45_0 + 1_0 + 16_{-1} + \overline{16}_1\,.$$

If the system were to contain any 27's of E_6 then it is clear that 2ν could only be an integer multiple of 3. However, our system contains only the adjoint representation. In effect our group is E_6/Z_3 and we conclude that 2ν may be any integer, $\nu = \frac{1}{2}n$. The field strength two-form $F = dA_\pm$ is then given by

$$F = \tfrac{1}{2}n\,Q\sin\theta\,d\theta \wedge d\varphi\,. \quad (4.5)$$

In an orthonormal basis the Ricci tensor reduces to

$$R_{\alpha\beta} = a^{-2}\delta_{\alpha\beta}\,, \quad \alpha,\beta = 4,5\,, \quad (4.6)$$

where a denotes the radius of S^2. In this basis the field strength (4.5) is represented by the component,

$$F^{Q}_{45} = n\tfrac{1}{2}a^{-2}\,. \quad (4.7)$$

On substituting (4.6) and (4.7) into (4.1) and (4.2) one obtains the algebraic equations

$$\kappa^{-2}a^{-2} = n^2\tfrac{1}{2}a^{-4}g_6^{-2}e^{\kappa\sigma}\,,$$

$$0 = n^2\tfrac{1}{4}a^{-4}g_6^{-2}e^{\kappa\sigma} - g_1^2\tfrac{1}{4}\kappa^{-4}e^{-\kappa\sigma}\,. \quad (4.8)$$

These equations fix the radius of S^2, $a^2 = 2\kappa^2 g_1^{-2}e^{\kappa\sigma}$, but they also imply a rather delicate constraint on the values of g_6 and g_1,

$$g_6 = \pm\tfrac{1}{2}ng_1\,. \quad (4.9)$$

Otherwise no solution of the desired type would be possible.

The fermionic zero modes are contained among the gauginos of E_6 since these are the only fields that couple to the monopole background. Indeed, according to (4.7) and $\nu = \frac{1}{2}n$ the coupling is of strength $\nu Q = \pm\frac{1}{2}n$ on the $\overline{16}_1$ and 16_{-1} contained in the 78. The four-dimensional massless chiral fermions therefore comprise $2|n|$ families of SO(10) and no antifamilies. They belong to the $|n|$-dimensional irreducible representation of the Kaluza–Klein SU(2) symmetry. These fermions are neutral with respect to E_7 but they carry the original U(1) charge, half the families are positive and the other half negative – so that no U(1) anomaly arises. The remaining symmetry, $SU(2) \times SO(10)$, is of course anomaly-free in four dimensions.

All other fermionic modes, including those coming from the gravitino are massive. Supersymmetry is broken in this compactification.

The U'(1) associated with the monopole background is broken. The mechanism explained by Witten [2] makes the gauge field massive but leaves the associated current conserved up to anomalies. Hence the four-dimensional theory should contain an axion.

In the boson sector the four-dimensional theory must contain the graviton and the gauge vectors corresponding to $SU(2) \times SO(10) \times E_7 \times U(1)$. The E_7 vectors do not seem to have any interesting physical role at present. They are "drones" pairing with the E_7 gauginos which were needed to cancel the gravitational anomaly. They should at least be confined, perhaps make glueballs, etc. The $SO(10) \times U(1)$ vectors couple to quarks and leptons but the SU(2) Kaluza–Klein

vectors will couple to these fermions only if $|n| > 1$. Unfortunately, we may be limited to $|n| = 1$ (two families) for reasons of classical stability [11] ‡5.

The question of classical stability needs to be further examined. We observe only that the massless scalar contained in σ will presumably develop a mass at the one-loop level. Along with this, the expectation value $\langle\sigma\rangle$, should be fixed, and hence the radius of the internal sphere.

Other open questions are: is the safe $E_6 \times E_7 \times U(1)$ in $D = 6$ somehow descended from the safe $E_8 \times E_8$ in $D = 10$? Are the counter-terms which were needed to attain gauge and gravitational symmetry, sufficient to compensate any supercurrent anomalies which may arise, or are further (gauge-invariant) counter-terms needed? Is the theory one-loop finite? Like the Green–Schwarz theory which is supposed to represent a limit of the superstring in ten dimensions, we may conjecture that our model relates to a string with internal symmetry or a membrane theory ‡6 in six dimensions [12]. This may motivate finiteness.

Appendix. The fermion terms are [5] ‡7

$$\begin{aligned}
\mathcal{L}_\psi &= \tfrac{1}{2}\bar\psi_\mu\gamma^{\mu\nu\rho}D_\nu\psi_\rho + \tfrac{1}{2}\bar\chi\gamma^\mu D_\mu\chi + \tfrac{1}{2}\bar\lambda^{\hat I}\gamma^\mu D_\mu\lambda^{\hat I} \\
&+ \tfrac{1}{2}\bar\psi^a\gamma^\mu D_\mu\psi_a + \tfrac{1}{2}\kappa\bar\psi_\mu\gamma^\nu\gamma^\mu\chi\partial_\nu\sigma \\
&- \kappa\bar\psi^A_\mu\gamma^\nu\gamma^\mu\psi^a(D_\nu\phi^\alpha)V_{\alpha aA} \\
&+ \tfrac{1}{24}\kappa^2 e^{\kappa\sigma}G_{\mu\nu\rho}(\bar\psi^\lambda\gamma_{[\lambda}\gamma^{\mu\nu\rho}\gamma_{\tau]}\psi^\tau + 2\bar\psi_\lambda\gamma^{\mu\nu\rho}\gamma^\lambda\chi \\
&- \bar\chi\gamma^{\mu\nu\rho}\chi + \bar\lambda^{\hat I}\gamma^{\mu\nu\rho}\lambda^{\hat I} + \bar\psi^a\gamma^{\mu\nu\rho}\psi_a) \\
&- 2^{-3/2}\kappa e^{\kappa\sigma/2}(\bar\psi_\lambda\gamma^{\mu\nu}\gamma^\lambda\lambda^{\hat I}F^{\hat I}_{\mu\nu} + \bar\chi\gamma^{\mu\nu}\lambda^{\hat I}F^{\hat I}_{\mu\nu}) \\
&+ 2^{-1/2}\kappa e^{-\kappa\sigma/2}[g_7\bar\psi_\mu\gamma^\mu T^i\lambda^I C^{iI}_7 + g_1\bar\psi_\mu\gamma^\mu T^i\lambda C^i_1 \\
&- g_7\bar\chi T^i\lambda^I C^{iI}_7 - g_1\bar\chi T^i\lambda C^i_1 \\
&- 2g_7\bar\psi^a\lambda^{IA}V_{\alpha aA}(T^I\phi)^\alpha \\
&- 2g_1\bar\psi^a\lambda^A V_{\alpha aA}(T^3\phi)^\alpha],
\end{aligned}$$

where the hyperscalars ϕ^α constitute a complex 912-vector of $Sp(456) \times U(1)$ charged with respect to $U(1)$. For inner products of $Sp(1)$ symplectic spinors, the contractions with ϵ_{AB} are always understood, e.g. $\bar\chi\gamma^{\mu\nu}\lambda^I = \bar\chi^A\gamma^{\mu\nu}\lambda^I_A$.

‡5 Other but less elegant anomaly-free models are: $E_6 \times Sp(1)$ with $81 = 78 + 3$ gauginos and 325 hyperino singlets; and $E_7 \times U(1)$ with $134 = 133 + 1$ gauginos plus $355 = 2 \times 133 + 2 \times 56$ hyperinos, 22 gaugino and 22 hyperino singlets.

‡6 We express appreciation to Dr. T. Jayaraman for pointing out ref. [12] to us.

‡7 We use the signature $(- + + + + +)$ here.

References

[1] M.B. Green and J. Schwarz, Phys. Lett. 149B (1984) 117.
[2] E. Witten, Phys. Lett. 149B (1984) 351.
[3] S. Randjbar Daemi, A. Salam and J. Strathdee, Phys. Lett. 124B (1983) 349;
G.F. Chapline and G.W. Gibbons, Phys. Lett. 135B (1984) 43.
[4] S. Okubo, J. Math. Phys. 23 (1982) 8.
[5] H. Nishino and E. Sezgin, Phys. Lett. 144B (1984) 187.
[6] J. Bagger and E. Witten, Nucl. Phys. B222 (1983) 1.
[7] I. Koh and H. Nishino, ICTP Trieste preprint IC/84/129 (1984).
[8] A. Salam and E. Sezgin, Phys. Lett. 147B (1984) 47.
[9] B. Zumino, Les Houches Lectures (1983), eds. R. Stora and B. De Witt (North-Holland, Amsterdam), to be published;
B. Zumino, Y.-S. Wu and A. Zee, Nucl. Phys. B239 (1984) 477;
W.A. Bardeen and B. Zumino, Berkeley preprint LBL17639 (1984).
[10] L. Alvarez Gaumé and E. Witten, Nucl. Phys. B234 (1984) 269.
[11] S. Randjbar-Daemi, A. Salam and J. Strathdee, Phys. Lett. 124B (1983) 345; Phys. Lett. 144B (1984) 455 (E);
A.N. Schellekens, Stability of higher dimensional Einstein–Yang–Mills theories on symmetric spaces, Stony Brook preprint ITP-SB-84-45 (1984).
[12] F. Ardalan and F. Mansouri, Phys. Rev. D9 (1974) 3341;
F. Mansouri and P.G.O. Freund, Z. Phys. C14 (1982) 279.

Nuclear Physics B269 (1986) 77–96
North-Holland, Amsterdam

LOCALLY SUPERSYMMETRIC σ-MODEL WITH WESS-ZUMINO TERM IN TWO DIMENSIONS AND CRITICAL DIMENSIONS FOR STRINGS

E. BERGSHOEFF

International Centre for Theoretical Physics, Trieste, Italy

S. RANDJBAR-DAEMI*

Institute for Theoretical Physics, Sidlerstr. 5, CH-3012 Bern, Switzerland

A. SALAM

International Centre for Theoretical Physics, Trieste, Italy and Imperial College, London, England

H. SARMADI

International Centre for Theoretical Physics, Trieste, Italy and International School of Advanced Studies, Trieste, Italy

E. SEZGIN

International Centre for Theoretical Physics, Trieste, Italy

Received 12 July 1985
(Revised 17 September 1985)

We construct an $N = 1$ locally supersymmetric σ-model with a Wess-Zumino term coupled to supergravity in two dimensions. If one takes the σ-model manifold to be the product of d-dimensional Minkowski space M^d and a group manifold G, and if the radius of G is quantized in appropriate units of the string tension, then the model describes a Neveu-Schwarz-Ramond (NSR)-type string moving on $M_d \times G$. (Our model generalizes earlier work of refs. [1,2] which do not contain a Wess-Zumino term and that of refs. [5,6] which is not locally supersymmetric.) The zweibein and the gravitino field equations yield constraints which generalize those of the NSR model to the case of a non-abelian group manifold. In particular, the fermionic constraint contains a new term trilinear in the fermionic fields. We quantize the theory in the light-cone gauge and derive the critical dimensions. We compute the mass spectrum of a closed string moving on $M_d \times G$ and show that massless fermions do not arise for non-abelian G for the spinning string, in agreement with the result of Friedan and Shenker [22].

* Supported by Swiss National Funds.

© Elsevier Science Publishers B.V.

1. Introduction

Some time ago, Deser and Zumino [1] and Brink, di Vecchia and Howe [2] constructed the coupling of ten scalar supersymmetric multiplets to $d = 2$ supergravity. They showed that the theory is conformally invariant and describes the $d = 10$ Neveu-Schwarz-Ramond [3] string model [15]. From the $d = 2$ point of view this theory is a locally supersymmetric σ-model in which the scalar manifold is a $d = 10$ flat Minkowski space-time, $M_{10} = ISO(9,1)/SO(9,1)$. In this paper we consider the generalization of this model in which the scalars parametrize an arbitrary riemannian manifold. In such a generalization, a Wess-Zumino term [4–6] is coupled to $d = 2$ supergravity. This model furnishes a covariant description of a string moving in curved space. In particular when this space is $M_d \times G$ where M_d is d-dimensional Minkowski space-time and G is a compact group manifold, the characteristic size of G is quantized in units of the string tension [7, 8b], while restrictions on d arise, due to the requirement of Lorentz invariance in M_d, of the quantized theory.

These results are relevant for the reduction of the critical dimensions in which the string theory can be consistently quantized. Recently, in refs. [7] and [8b] it was shown that the critical dimension for the bosonic string is given by

$$d = 26 - \frac{d_G}{1 + c_A/2|k|} \qquad \text{(bosonic string)}, \tag{1.1}$$

where d_G is the dimension of the group G, k is an integer and c_A is the eigenvalue of the second Casimir operator of G in the adjoint representation (see eq. (1.3)). Extending this result, in ref. [7] it was conjectured that the critical dimension for the fermionic string is given by

$$d = 10 - \frac{2}{3}\frac{d_G}{1 + c_A/2|k|} - \tfrac{1}{3}d_G \qquad \text{(spinning string)}. \tag{1.2}$$

The value of c_A for Lie groups is given by

G	SU(n)	SO($2n+1$)	Sp(n)	SO($2n$)	G_2	F_4	E_6	E_7	E_8
c_A	$2n$	$4n-2$	$2n+2$	$4n-4$	8	18	24	36	60

. (1.3)

In this paper, starting from our locally supersymmetric action (see eq. (2.1)) we quantize the theory in the light-cone gauge, rederive (1.1) and verify (1.2). We emphasize that the main body of this paper deals with (i) the construction of the $N = 1$, $d = 2$ locally supersymmetric σ-model with Wess-Zumino term and (ii) the derivation of the critical dimension formula for the fermionic superstring based on this model.

The case of $k = 1$ is special, since, as was shown by Witten [9], in this case the scalars of the non-linear σ-model become equivalent to free fermions. For $k = 1$, the

solution to (1.1) includes $G = SU(n)$, $SO(2n)$, E_6, E_7, E_8 or any product of these groups with*

$$r = 26 - d, \tag{1.4}$$

where r is the rank of G**. As is well known, for $d = 10$, $E_8 \times E_8$ is a solution of (1.4) which was used in the construction of the heterotic string [12].

For $k = 1$, assuming that G is simple, the unique solution to (1.2) is***

$$\begin{aligned} &d = 8, \quad G = SO(3), \\ &d = 6, \quad G = SU(3), \\ &\qquad\qquad\qquad\quad \text{(spinning string)} \\ &d = 5, \quad G = SO(5), \\ &d = 3, \quad G = SU(4). \end{aligned} \tag{1.5}$$

(Note that $d = 4$ is not included for any *simple* G.) All these solutions refer to classically and quantum mechanically consistent *free* string theories. At present it is not clear how to formulate consistent interacting string theories based on these solutions.

This paper is organized as follows. In sect. 2 we construct the $N = 1$ locally supersymmetric action with a Wess-Zumino term. In sect. 3 we derive the field equations and constraints following from the action. In sect. 4 we quantize the system and derive the critical dimension formula by the requirement of Lorentz invariance in M_d. In sect. 5 we discuss the spectrum of a closed string moving on $M_d \times G$. For the bosonic case it coincides with that of the string which compactifies on r-tori where r is the rank of a simply laced group. Various aspects of this phenomenon have been discussed by several authors [7, 8a, 19, 20]. In the fermionic case, we find that massless fermions do not arise for nonabelian groups, in agreement with the result of Friedan and Shenker [22]. Finally in sect. 6 we discuss some of the open problems.

* For $k > 1$, some of the solutions to (1.1) are: $d = 7$, E_7 ($k = 3$); $d = 8$, SU(5) ($k = 15$); $d = 6$, SU(7) ($k = 5$).

** For a complete discussion see Goddard, Nahm and Olive (and other papers) in ref. [8a]. (It has recently been shown [10] that for arbitrary k the $O(N)$ σ-model is equivalent to free particles obeying parastatistics of order k. Regarding the question of parastatistics in strings see also ref. [11] where it is shown that $d = 2 + 8/q$ for a free fermionic theory exhibiting parastatistics of order q. For $q = 2$ this gives $d = 6$.)

*** For $k > 1$, the unique solution (for simple G) is: $d = 8$, SU(2) ($k = 2$); $d = 4$, SO(5) ($k = 2$), SU(3) ($k = 5$); $d = 2$, SO(5) ($k = 7$).

2. A locally supersymmetric action with the Wess-Zumino term

We construct the action of an $N = 1$ locally supersymmetric σ-model coupled to supergravity in two dimensions, where the scalars of the σ-model parametrize an arbitrary riemannian manifold M. $N = 1$ supergravity in two dimensions contains a zweibein e_μ^a and a gravitino ψ_μ. As is well known, in two dimensions these fields do not describe physical degrees of freedom. Nevertheless they play an important role in that their field equations yield constraints on the scalars and spinors of the σ-model.

The full action reads as follows:

$$
\begin{aligned}
e^{-1}\mathcal{L} = \frac{1}{2\pi\alpha'}\Big[& -\tfrac{1}{2}g^{\mu\nu}\,\partial_\mu\phi^i\,\partial_\nu\phi^j\,g_{ij} - \tfrac{1}{2}i\bar{\chi}^i\gamma^\mu\big(\partial_\mu\chi^j + \Gamma^j_{kl}\,\partial_\mu\phi^k\,\chi^l\big)g_{ij} \\
& + \bar{\psi}_\mu\gamma^\nu\gamma^\mu\chi^i\,\partial_\nu\phi^j\,g_{ij} - \tfrac{1}{4}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\psi_\nu\bar{\chi}^i\chi^j g_{ij} - \tfrac{1}{12}R_{ijkl}\bar{\chi}^i\chi^k\bar{\chi}^j\chi^l \\
& - e^{-1}\frac{k}{8\pi}\varepsilon^{\mu\nu}b_{ij}\,\partial_\mu\phi^i\,\partial_\nu\phi^j - \frac{ik}{16\pi}T_{ijk}\bar{\chi}^i\gamma^\mu\gamma_5\chi^j\,\partial_\mu\phi^k \\
& + \frac{k}{64\pi}T_{il[j;k]}\bar{\chi}^i\chi^k\bar{\chi}^j\gamma_5\chi^l - \frac{k^2}{512\pi^2}g^{mn}T_{ikm}T_{jln}\bar{\chi}^i\gamma_5\chi^k\bar{\chi}^j\gamma_5\chi^l \\
& + \frac{ik}{48\pi}T_{ijk}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\chi^i\bar{\chi}^j\gamma_\nu\gamma_5\chi^k\Big],
\end{aligned}
$$

$$
\mu = 0,1, \qquad i = 1,\ldots,\dim \mathrm{M} \tag{2.1}
$$

and is invariant under the following transformations:

$$
\begin{aligned}
\delta e_\mu^a &= 2i\bar{\varepsilon}\gamma^a\psi_\mu - \Lambda e_\mu^a, \\
\delta\psi_\mu &= \big(\partial_\mu - \tfrac{1}{4}\omega_\mu^{ab}\gamma_{ab}\big)\varepsilon + i\gamma_\mu\eta - \tfrac{1}{2}\Lambda\psi_\mu, \\
\delta\phi^i &= -\bar{\varepsilon}\chi^i, \\
\delta\chi^i &= -i\gamma^\mu\big(\partial_\mu\phi^i + \bar{\psi}_\mu\chi^i\big)\varepsilon - \Gamma^i_{jk}\delta\phi^j\chi^k + \frac{k}{16\pi}g^{il}T_{ljk}\big(\bar{\chi}^j\gamma_5\chi^k\big)\varepsilon + \tfrac{1}{2}\Lambda\chi^i,
\end{aligned} \tag{2.2}
$$

where $\varepsilon(\sigma,\tau)$, $\eta(\sigma,\tau)$ and $\Lambda(\sigma,\tau)$ are the supersymmetry, conformal supersymmetry and Weyl scale transformation parameters, respectively. In (2.1) and (2.2) we have used the following definitions and conventions. The scalars ϕ^i $(i = 1,\ldots\dim \mathrm{M})$ parametrize a riemannian manifold, M, with metric $g_{ij}(\phi)$. The spinors χ^i and ψ_μ

are two-component Majorana*. The parameter α' is the coupling constant of the σ-model and the coefficient k in front of the Wess-Zumino term is defined such that the path integral corresponding to (2.1) is well defined [9] (and last reference in [6]). Thus k is an integer. Our Riemann tensor is defined as $R^i_{jkl} = \partial_l \Gamma^i_{jk} + \cdots$. Note that $\gamma^\mu = \gamma^a e^\mu_a(\mu, a = 0, 1)$. The 3-form T_{ijk} is closed and is the curl of a second-rank antisymmetric tensor $b_{ij}(\phi) = -b_{ji}(\phi)$, which is a function of the scalars ϕ:

$$T_{ijk} = \partial_i b_{jk} + \partial_k b_{ij} + \partial_j b_{ki}. \tag{2.3}$$

The spin connection, ω^{ab}_μ, contains the contorsion tensor $K_{\mu ab} = 2i\bar\psi_a \psi^\mu \psi_b$. The term in $\delta\chi^i$ containing the Christoffel symbol Γ^i_{jk} has been added so that when it is taken to the left-hand side, $(\delta\chi^i + \Gamma^i_{jk}\delta\phi^j\chi^k)$ transforms as a vector on M, as it should since the remaining terms on the right-hand side have the same transformation property. Finally, the notation ; is standard for riemannian covariant derivatives.

In order to derive (2.1) and (2.2) we proceed in two steps. First we generalize the action and transformation rules given in refs. [1] and [2] to the case of an arbitrary riemannian manifold, M, with metric $g_{ij}(\phi)$, as follows:

$$e^{-1}\mathcal{L}_1 = \frac{1}{2\pi\alpha'}\Big[-\tfrac{1}{2}g^{\mu\nu}\partial_\mu\phi^i\partial_\nu\phi^j g_{ij} - \tfrac{1}{2}i\bar\chi^i\gamma^\mu\big(\partial_\mu\chi^j + \Gamma^k_{kl}\partial_\mu\phi^k\chi^l\big)g_{ij}$$

$$+\bar\psi_\mu\gamma^\nu\gamma^\mu\chi^i\partial_\nu\phi^j g_{ij} - \tfrac{1}{4}\bar\psi_\mu\gamma^\nu\gamma^\mu\psi_\nu\bar\chi^I\chi^j g_{ij} - \tfrac{1}{12}R_{ijkl}\bar\chi^i\chi^k\bar\chi^j\chi^l\Big] \tag{2.4}$$

and

$$\delta e^a_\mu = 2i\bar\varepsilon\gamma^a\psi_\mu - \Lambda e^a_\mu,$$

$$\delta\psi_\mu = \big(\partial_\mu - \tfrac{1}{4}\omega^{ab}_\mu\gamma_{ab}\big)\varepsilon + i\gamma_\mu\eta - \tfrac{1}{2}\Lambda\psi_\mu,$$

$$\delta\phi^i = -\bar\varepsilon\chi^i,$$

$$\delta\chi^i = -i\gamma^\mu\big(\partial_\mu\phi^i + \bar\psi_\mu\chi^i\big)\varepsilon - \Gamma^i_{jk}\delta\phi^j\chi^k + \tfrac{1}{2}\Lambda\chi^i. \tag{2.5}$$

Remarkably we find that without any further modifications the covariantized action (2.4) is already fully invariant under the local supersymmetry and scale transformations given in (2.5). In fact a large class of variations are those which arise in either the model of refs. [1] and [2] ($g_{\mu\nu} \neq \eta_{\mu\nu}$, $g_{ij} = \eta_{ij}$) provided that in the

* We use the following conventions: Our metric is $\eta_{ab} = \mathrm{diag}(-1, +1)$. The $d = 2$ gamma matrices γ^a ($a = 0, 1$) are $\gamma^0 = \sigma_2$, $\gamma^1 = i\sigma_1$ and $\gamma_5 = -\sigma_3$. We take $\varepsilon^{01} = -1$, $\varepsilon_{01} = +1$ and $\bar\chi = \chi^T\gamma^0$. In two dimensions the following Fierz relations hold: $\psi\bar\lambda = -\frac{1}{2}(\bar\lambda\psi + (\bar\lambda\gamma_5\psi)\gamma_5 - (\bar\lambda\gamma_\mu\psi)\gamma_\mu)$ and $\psi(\bar\lambda\chi) = -\lambda(\bar\chi\psi) - \chi(\bar\psi\lambda)$. Some further useful relations are: $\gamma_\mu\gamma_\nu = -g_{\mu\nu} + e^{-1}\varepsilon_{\mu\nu}\gamma_5$, $e = \det e^a_\mu$ and $\varepsilon^{\mu\nu}\gamma_\nu = \gamma^\mu\gamma_5$.

variation of the action the derivatives are covariantized with respect to M (e.g. $\partial_\mu \chi^i \to \partial_\mu \chi^i + \Gamma^i_{jk} \partial_\mu \phi^j \chi^k$) or the globally supersymmetric σ-model [13] ($g_{\mu\nu} = \eta_{\mu\nu}$, $g_{ij} \neq \eta_{ij}$). Therefore one only has to check the cancellation of the new variations which do not fall into these two classes. Most cancellations are trivial, the only non-trivial ones being those arising from the variation of the zweibein and χ^i in the $R\chi^4$ term. These variations give a vanishing result:

$$-\tfrac{1}{6} i\bar{\varepsilon}\gamma^\lambda \psi_\lambda R_{ijkl} \bar{\chi}^i \chi^k \bar{\chi}^j \chi^l + \tfrac{1}{6} R_{ijkl} \bar{\chi}^i \chi^k \bar{\chi}^j \big(i\gamma^\nu \varepsilon \bar{\psi}_\nu \chi^l\big)$$

$$+ \tfrac{1}{6} R_{ijkl} \bar{\chi}^i \chi^k \bar{\chi}^l \big(i\gamma^\nu \varepsilon \bar{\psi}_\nu \chi^j\big) = 0. \tag{2.6}$$

To prove this, one must Fierz rearrange χ^j and χ^l in the second and third term and use the fact that

$$R_{ijkl} \bar{\chi}^i \chi^k \bar{\chi}^j \gamma_5 \chi^l = 0 = R_{ijkl} \bar{\chi}^i \chi^k \bar{\chi}^j \gamma_\mu \chi^l. \tag{2.7}$$

This identity is easily proven by noting that R_{ijkl} is symmetric, whereas the χ terms are antisymmetric in the pair interchange $(ij) \leftrightarrow (kl)$. We thus conclude that the action given in (2.4) is invariant under (2.5).

We now consider the extension of (2.4), (2.5) by adding a Wess-Zumino term in a locally supersymmetric manner. A globally supersymmetric Wess-Zumino term has been constructed in refs. [5,6]. In the case of local supersymmetry we leave the transformation rules (2.5) intact except for $\delta\chi^i$ which we modify to read

$$\delta\chi^i = -i\gamma^\mu \big(\partial_\mu \phi^i + \bar{\psi}_\mu \chi^i\big)\varepsilon - \Gamma^i_{jk} \delta\phi^j \chi^k + \frac{k}{16\pi} g^{il} T_{ljk} \big(\bar{\chi}^j \gamma_5 \chi^k\big)\varepsilon. \tag{2.8}$$

Note that the k-dependent term in $\delta\chi^i$ is precisely the one which occurs in the globally supersymmetric model. To obtain an action which is invariant under the local supersymmetries (2.5), (2.8), we first covariantize the globally supersymmetric Wess-Zumino action [5,6] with respect to the two-dimensional space-time. This yields

$$e^{-1}\mathcal{L}_2 = -e^{-1} \frac{k}{8\pi} \varepsilon^{\mu\nu} b_{ij}\, \partial_\mu \phi^i\, \partial_\nu \phi^j - \frac{ik}{16\pi} T_{ijk} \bar{\chi}^i \gamma^a \gamma^5 \chi^j\, \partial_\mu \phi^k\, e^\mu_a$$

$$+ \frac{k}{64\pi} T_{ilj;k} \bar{\chi}^i \chi^k \bar{\chi}^j \gamma_5 \chi^l - \frac{k^2}{512\pi^2} g^{mn} T_{ikm} T_{jln} \bar{\chi}^i \gamma_5 \chi^k \bar{\chi}^j \gamma_5 \chi^l. \tag{2.9}$$

We now consider the new variations in $\mathcal{L}_1 + \mathcal{L}_2$, given in (2.4) and (2.9), which do not arise in the globally supersymmetric Wess-Zumino term of refs. [5] and [6]. One class of variations are the terms proportional to T^2 coming from the variation of the

determinant e and of χ^i in the $T^2\chi^4$ term. These variations are

$$-\frac{1}{2\pi\alpha'}\left[\frac{k^2}{512\pi^2}T_{ikm}T_{jl}{}^m\bar{\chi}^i\gamma_5\chi^k\bar{\chi}^j\gamma_5\chi^l\left(2i\bar{\varepsilon}\gamma^\lambda\psi_\lambda\right)\right.$$

$$\left.+\frac{k^2}{128\pi^2}T_{ikm}T_{jl}{}^m\bar{\chi}^i\gamma_5\chi^k\bar{\chi}^j\gamma_5\left(-i\gamma^\nu\varepsilon\bar{\psi}_\nu\chi^l\right)\right]. \tag{2.10}$$

The Fierz rearrangement of $\chi^l\bar{\chi}^j$ in the second term yields two terms one of which cancels the first term. The final result is

$$\frac{-ik^2}{512\alpha'\pi^3}T_{ikm}T_{jl}{}^m\bar{\chi}^i\gamma_5\chi^k\bar{\chi}^j\gamma_\lambda\chi^l\bar{\varepsilon}\gamma^\nu\gamma_5\gamma^\lambda\psi_\nu. \tag{2.11}$$

In order to cancel this we add to the action the following new term:

$$e^{-1}\mathcal{L}_3=\frac{ik}{96\pi^2\alpha'}T_{ijk}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\chi^i\bar{\chi}^j\gamma_\nu\gamma_5\chi^k. \tag{2.12}$$

One can easily show by appropriate Fierz rearrangements that all the χ's in (2.12) contribute with the same weight to $\delta\mathcal{L}_3$. In particular, the T-dependent variation of the χ's in (2.12) gives

$$3\times\frac{ik}{96\pi^2\alpha'}\left(T_{ijk}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\right)\left(\frac{k}{16\pi}T_{imn}\bar{\chi}^m\gamma_5\chi^n\bar{\chi}^j\gamma_\nu\gamma_5\chi^k\varepsilon\right), \tag{2.13}$$

which exactly cancels (2.11).

Remarkably we now find that with the addition of $\mathcal{L}_3$ given in (2.12) the full lagrangian $\mathcal{L}=\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3$ given in (2.4), (2.9), (2.12) or (2.1) is invariant under the local supersymmetries (2.5), (2.8) or (2.2) without any further modifications. We have checked this by a straightforward but tedious calculation which will not be reproduced.

3. Field equations and constraints

We now study the field equations which follow from the action, (2.1). As mentioned in the introduction the scalar manifold is taken to be $M_d\times G$. We shall see that for a particular relation among the parameters of the theory (see eq. (3.7), the field equations for ϕ^i and χ^i will be completely integrable. The field equations for e_μ^a and ψ_μ lead to constraints on ϕ^i and χ^i which generalize those of the Neveu-Schwarz-Ramond model. These constraints will play an important role in determining the critical dimension for strings in curved space, as we shall show in the next section.

We first choose co-ordinates appropriate to the product structure $M_d \times G$:

$$\phi^i = \begin{pmatrix} x^\alpha \\ y^I \end{pmatrix}, \qquad \chi^i = \begin{pmatrix} \lambda^\alpha \\ \chi^I \end{pmatrix}, \qquad \alpha = 0,1,\ldots,d-1, \qquad I = 1,\ldots,d_G. \quad (3.1a)$$

Correspondingly

$$g_{ij} = \begin{pmatrix} \eta_{\alpha\beta} & 0 \\ 0 & g_{IJ}(y) \end{pmatrix}, \qquad b_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{2\pi\alpha'}{R^2} b_{IJ}(y) \end{pmatrix}, \quad (3.1b)$$

where $\eta_{\alpha\beta}$ is the usual metric in M_d and g_{IJ} is the $G \times G$ invariant metric on G. Note that we have introduced an additional parameter R, which is associated with the characteristic size of the compact group manifold G. In terms of the coordinates x^α and y^I the lagrangian (2.1) splits into two parts which are separately invariant. One part, $\mathcal{L}(M_d)$, depends only on x^α and has exactly the same form as the lagrangian given in refs. [1] and [2]

$$e^{-1}\mathcal{L}(M_d) = \frac{1}{2\pi\alpha'}\Big[-\tfrac{1}{2}g^{\mu\nu}\,\partial_\mu x^\alpha\,\partial_\nu x^\beta\,\eta_{\alpha\beta} - \tfrac{1}{2}i\bar{\lambda}^\alpha\gamma^\mu\,\partial_\mu\lambda^\beta\,\eta_{\alpha\beta}$$
$$+\bar{\psi}_\mu\gamma^\nu\gamma^\mu\lambda^\alpha\big(\partial_\nu x^\beta\big)\eta_{\alpha\beta} - \tfrac{1}{4}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\psi_\nu\bar{\lambda}^\alpha\lambda^\beta\eta_{\alpha\beta}\Big]. \quad (3.2)$$

The other part $\mathcal{L}(G)$ depends only on the group co-ordinates y^I. On a group manifold G we can take T_{IJK} to be

$$T_{IJK} = \frac{1}{R}f_{abc}L_I^a L_J^b L_K^c = -\frac{1}{R}f_{abc}R_I^a R_J^b R_K^c \equiv \frac{1}{R}f_{IJK}, \quad (3.3)$$

where $f_{ab}{}^c$ are the structure constants of the group G and $L_I^a(\phi)$ are the left-invariant basis elements on the group manifold G, which are defined by

$$g^{-1}\,\partial_\mu g = \frac{1}{R}L_I^a T_a\,\partial_\mu y^I, \qquad g\,\partial_\mu g^{-1} = -\frac{1}{R}R_I^a T_a\,\partial_\mu y^I. \quad (3.4)$$

Here g is a group element and T^a are the antihermitian generators of the Lie algebra of G. From (3.3) it follows that T_{IJK} is an invariant 3-form on G and hence that b_{IJ} is an invariant 2-form (up to a total derivative).

Using (3.3) and the following relations [14, 13]

$$R_{IJKL} = \frac{-1}{4R^2}f_{IJ}{}^M f_{KLM}, \quad (3.5a)$$

$$R_{IJKL}\bar{\chi}^I\chi^K\bar{\chi}^J\chi^L = 3R_{IJKL}\bar{\chi}^I\gamma_5\chi^K\bar{\chi}^J\gamma_5\chi^L, \quad (3.5b)$$

the Lagrangian $\mathcal{L}(G)$ now reads

$$
\begin{aligned}
e^{-1}\mathcal{L}(G) = {} & \frac{-1}{4\pi\alpha'}\,\partial_\mu y^I\,\partial_\nu y^J\left(g^{\mu\nu}g_{IJ}+\frac{k\alpha'}{2R^2}e^{-1}\varepsilon^{\mu\nu}b_{IJ}\right)\\
& -\frac{i}{4\pi\alpha'}\bar{\chi}^I\gamma^\mu\left[\partial_\mu\chi^J+\left(\Gamma^J_{KL}-\frac{k\alpha'}{4R^3}\gamma_5 f^J_{KL}\right)\partial_\mu y^K\chi^L\right]g_{IJ}\\
& +\frac{1}{32\pi\alpha' R^2}f_{IJ}{}^M f_{KLM}\left(1-\frac{\alpha'^2k^2}{4R^4}\right)\bar{\chi}^I\gamma_5\chi^K\bar{\chi}^J\gamma_5\chi^L\\
& +\frac{1}{2\pi\alpha'}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\chi^I\,\partial_\nu y^J g_{IJ}-\frac{1}{8\pi\alpha'}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\psi_\nu\bar{\chi}^I\chi^J g_{IJ}\\
& +\frac{ik}{48\pi R^3}f_{IJK}\bar{\psi}_\mu\gamma^\nu\gamma^\mu\chi^I\bar{\chi}^J\gamma_\nu\gamma_5\chi^K. \qquad (3.6)
\end{aligned}
$$

We see that for

$$R^2=\tfrac{1}{2}|k|\alpha' \qquad (3.7)$$

the quartic χ^4 terms in (3.6) cancel. Without loss of generality we shall always take k to be positive. Using (3.7), in the superconformal gauge

$$e^a_\mu = f(\sigma,\tau)\,\delta^a_\mu, \qquad \psi_\mu=\gamma_\mu\lambda(\sigma,\tau) \qquad (3.7')$$

the field equations following from (3.2) and (3.6) are

$$A_1=\eta^{\mu\nu}\,\partial_\mu\partial_\nu x^\alpha=0, \qquad A_2=\gamma^\mu\,\partial_\mu\lambda^\alpha=0, \qquad (3.8a,b)$$

$$A_3=\eta^{\mu\nu}\,\partial_\mu\partial_\nu y^I+\partial_\mu y^J\,\partial_\nu y^k\left(g^{\mu\nu}\Gamma^I_{JK}-\frac{1}{2R}\varepsilon^{\mu\nu}f_{JK}{}^I\right)=0, \qquad (3.8c)$$

$$A_4=\gamma^\mu\,\partial_\mu\chi^I+\gamma^\mu\left(\Gamma^I_{JK}-\frac{1}{2R}f_{JK}{}^I\gamma_5\right)\partial_\mu y^J\chi^K=0, \qquad (3.8d)$$

$$A_5=\gamma^\nu\gamma^\mu\lambda_\alpha\,\partial_\nu x^\alpha+\gamma^\nu\gamma^\mu\chi^I\,\partial_\nu y^J g_{IJ}+\frac{i}{12R}f_{IJK}\gamma^\nu\gamma^\mu\chi^I\bar{\chi}^J\gamma_\nu\gamma_5\gamma^K=0, \qquad (3.8e)$$

$$
\begin{aligned}
A_6 = {} & \eta_{\alpha\beta}\left(-\partial_\mu x^\alpha\,\partial_\nu x^\beta+\tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\,\partial_\rho x^\alpha\,\partial_\sigma x^\beta\right)-\tfrac{1}{2}i\bar{\lambda}^\alpha\gamma_\nu\,\partial_\mu\lambda_\alpha\\
& +g_{IJ}\left(-\partial_\mu y^I\,\partial_\nu y^J+\tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\,\partial_\rho y^I\,\partial_\sigma y^J\right)-\tfrac{1}{2}i\bar{\chi}^I\gamma_\nu\,\partial_\mu\chi^J g_{IJ}\\
& -\tfrac{1}{2}i\bar{\chi}^I\gamma_\nu\left(\Gamma^J_{MN}-\frac{1}{2R}\gamma_5 f_{MN}{}^J\right)\partial_\mu y^M\chi^N g_{IJ}=0. \qquad (3.8f)
\end{aligned}
$$

Note that $A_1 = A_2 = A_3 = A_4 = 0$ are the physical field equations, while $A_5 = A_6 = 0$ are constraint equations. The field equation $A_3 = 0$ corresponds to the variation $\delta I/\delta y^I - \Gamma^K_{IJ}(\delta I/\delta \chi^K)\chi^J = 0$ to ensure covariance. This variation gives the result (3.8c) plus a term which vanishes upon use of the χ^I field equation. Note that the resulting equation, $A_3 = 0$, does not depend on χ^I. Also in (3.8f) the χ^I field equation has been used.

It is convenient to rewrite (3.8c) in terms of the bosonic part of the currents associated with the right and left translations on the group manifold. These currents are defined by

$$J^{\mu a}_{\mathrm{R}} = \tfrac{1}{2}(\eta^{\mu\nu} + \varepsilon^{\mu\nu})\,\partial_\nu y^I L^a_I, \tag{3.9a}$$

$$J^{\mu a}_{\mathrm{L}} = \tfrac{1}{2}(\eta^{\mu\nu} - \varepsilon^{\mu\nu})\,\partial_\nu y^I R^a_I. \tag{3.9b}$$

Furthermore, the following relations will be useful in simplifying the field equations [14]

$$g_{IJ} = L^a_I L^a_J = R^a_I R^a_J,$$

$$\Gamma^I_{JK} = -L^a_K \partial_J L^I_a + \frac{1}{2R} f_{JK}{}^I = -R^a_K \partial_J R^I_a - \frac{1}{2R} f_{JK}{}^I,$$

$$\partial_I L^a_J - \partial_J L^a_I + \frac{1}{R} f_{IJ}{}^K L^a_K = 0 = \partial_I R^a_J - \partial_J R^a_I - \frac{1}{R} f_{IJ}{}^K R^a_K. \tag{3.10}$$

Using (3.9) and (3.10), we can rewrite (3.8c) as

$$\partial_\mu J^{\mu a}_{\mathrm{R}} = 0, \qquad \partial_\mu J^{\mu a}_{\mathrm{L}} = 0. \tag{3.11a,b}$$

To simplify $A_4 = A_5 = A_6 = 0$ we now define the components of χ^I in the left and right invariant basis elements as follows:

$$\chi^a = \chi^I L^a_I, \qquad \tilde{\chi}^a = \chi^I R^a_I. \tag{3.12}$$

It is not difficult to show that the supersymmetry variation of $J^{\mu a}_{\mathrm{R}}$ contains $\chi^a_{\mathrm{L}} = \frac{1}{2}(1 + \gamma_5)\chi^a$ and that of $J^{\mu a}_{\mathrm{L}}$ contains $\tilde{\chi}^a_{\mathrm{R}} = \frac{1}{2}(1 - \gamma_5)\tilde{\chi}^a$. Thus it is natural to express equations $A_4 = A_5 = A_6 = 0$ in terms of χ^a_{L} and $\tilde{\chi}^a_{\mathrm{R}}$. Again using (3.9) and (3.10), from (3.8d, e, f) it now follows that

$$\gamma^\mu \partial_\mu \chi^a_{\mathrm{L}} = 0, \qquad \gamma^\mu \partial_\mu \tilde{\chi}^a_{\mathrm{R}} = 0, \tag{3.13a,b}$$

$$(\eta^{\mu\nu} + \varepsilon^{\mu\nu})\left(\lambda_{\mathrm{L}\alpha}\,\partial_\nu x^\alpha + \chi^a_{\mathrm{L}} J^a_{\nu\mathrm{R}} - \frac{i}{12R} f_{abc} \chi^a_{\mathrm{L}} \bar{\chi}^b_{\mathrm{L}} \gamma_\nu \chi^c_{\mathrm{L}}\right) = 0, \tag{3.14a}$$

$$(\eta^{\mu\nu} - \varepsilon^{\mu\nu})\left(\lambda_{\mathrm{R}\alpha}\,\partial_\nu x^\alpha + \tilde{\chi}^a_{\mathrm{R}} J^a_{\nu\mathrm{L}} - \frac{i}{12R} f_{abc} \tilde{\chi}^a_{\mathrm{R}} \bar{\tilde{\chi}}^b_{\mathrm{R}} \gamma_\nu \tilde{\chi}^c_{\mathrm{R}}\right) = 0, \tag{3.14b}$$

$$\begin{aligned}
&\eta_{\alpha\beta}\left(-\partial_\mu x^\alpha\,\partial_\nu x^\beta + \tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\,\partial_\rho x^\alpha\,\partial_\sigma x^\beta\right) - \tfrac{1}{2} i \bar{\lambda}^\alpha_{\mathrm{L}} \gamma_\nu\,\partial_\mu \lambda_{\mathrm{L}\alpha} \\
&\quad - \tfrac{1}{2} i \bar{\lambda}^\alpha_{\mathrm{R}} \gamma_\nu\,\partial_\mu \lambda_{\mathrm{R}\alpha} + \eta_{\alpha\beta}\left(-\partial_\mu y^I\,\partial_\nu y^J + \tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\partial_\rho y^I\,\partial_\sigma y^J\right) \\
&\quad - \tfrac{1}{2} i \bar{\chi}^a_{\mathrm{L}} \gamma_\nu\,\partial_\mu \chi^a_{\mathrm{L}} - \tfrac{1}{2} i \bar{\tilde{\chi}}^a_{\mathrm{R}} \gamma_\nu\,\partial_\mu \tilde{\chi}^a_{\mathrm{R}} = 0.
\end{aligned} \tag{3.15}$$

Note that the last terms in (3.8d, f) have dropped out, due to the fact that we have expressed $\chi^I_{L,R}$ in the appropriate basis.

In summary, the field equations following from the action (3.2), (3.6) are given by (3.8a, b), (3.11a, b) and (3.13a, b), while the constraints are given by (3.14) and (3.15). In particular, note the presence of the χ^3 terms in (3.14) which are necessary for the closure of the super-Virasoro algebra. We will use these equations as a starting point for the quantization of the system in the next section.

4. The critical dimension formula for strings moving on $M^d \times G$

We now solve the free field equations given in (3.8), (3.11) and (3.13). Next we quantize the system subject to the constraints given in (3.14) and (3.15). The requirement of Lorentz invariance of the quantized theory in M_d puts restrictions on the dimensions d and d_G of M_d and the group manifold G, respectively. These restrictions have recently been obtained for the bosonic string in refs. [7] and [8b] and conjectured for the fermionic string in ref. [7]. In this section we shall restrict ourselves to closed strings. The case of open string can be treated similarly.

The solutions to the field equations are given by (setting $2\alpha' = 1$)*

$$x^\alpha(\tau,\sigma) = q^\alpha + p^\alpha\tau + \tfrac{1}{2}i\sum_{n\neq 0}\frac{1}{n}\left(\alpha_n^\alpha e^{-2in(\tau+\sigma)} + \tilde{\alpha}_n^\alpha e^{-2in(\tau-\sigma)}\right), \tag{4.1a}$$

$$J^a_{+R} = \sqrt{2}\sum_{n=-\infty}^{\infty}{}' \beta_n^a e^{-2in(\tau+\sigma)}, \tag{4.1b}$$

$$J^a_{-L} = \sqrt{2}\sum_{n=-\infty}^{\infty} \tilde{\beta}_n^a e^{-2in(\tau-\sigma)}, \tag{4.1c}$$

$$\lambda^{\alpha(1)} = \sum_{n=-\infty}^{\infty} \tilde{d}_n^\alpha e^{-2in(\tau-\sigma)}, \tag{4.1d}$$

$$\lambda^{\alpha(2)} = \sum_{n=-\infty}^{\infty} \tilde{d}_n^\alpha e^{-2in(\tau+\sigma)}, \tag{4.1e}$$

$$\chi^{a(1)} = \sum_{n=-\infty}^{\infty} S_n^a e^{-2in(\tau-\sigma)}, \tag{4.1f}$$

$$\chi^{a(2)} = \sum_{n=-\infty}^{\infty} \tilde{S}_n^a e^{-2in(\tau+\sigma)}, \tag{4.1g}$$

* The light-cone coordinates on the string world sheet are defined by $\xi^\pm = \sqrt{\frac{1}{2}}(\tau \pm \sigma)$, and those on M_d by $x^\pm = \sqrt{\frac{1}{2}}(x^0 \pm x^{d-1})$. In this section the index i refers to transverse directions on M_d.

where $\lambda^{\alpha(1)}(\lambda^{\alpha(2)})$ is the single non-vanishing component of $\lambda^{\alpha}_{L}(\lambda^{\alpha}_{R})$. In (4.1d) and (4.1e) the sums are over integers (half integers) if a periodic (antiperiodic) boundary condition for λ is chosen, and similarly for (4.1f) and (4.1g). To eliminate the gauge degrees of freedom, we use the light-cone gauge defined by

$$x^{+} = P^{+}\tau, \qquad \lambda^{+(1)} = 0, \qquad \lambda^{+(2)} = 0. \tag{4.2}$$

In this gauge, substituting (4.1) into the constraint equations (3.14) and $++$ and $--$ projections of (3.15), we solve for the Fourier components of x^{-} and λ^{-}, respectively, as follows:

$$\alpha_n^- = \frac{2}{p^+}\left(\tfrac{1}{2}\sum_{m=-\infty}^{\infty}\alpha^i_{n-m}\alpha^i_m + \tfrac{1}{2}\sum_{m=-\infty}^{\infty}\beta^a_{n-m}\beta^a_m \right.$$

$$\left. + \tfrac{1}{2}\sum_{m=-\infty}^{\infty}\left(m-\tfrac{1}{2}n\right)d^i_{n-m}d^i_m + \tfrac{1}{2}\sum_{m=-\infty}^{\infty}\left(m-\tfrac{1}{2}n\right)S^a_{n-m}S^a_m\right) \equiv \frac{2}{p^+}L_n, \tag{4.3a}$$

$$\tilde{\alpha}_n^- = \frac{2}{p^+}\tilde{L}_n, \qquad p^- = 2\alpha_0^- = 2\tilde{\alpha}_0^-, \qquad p^i = 2\alpha_0^i = 2\tilde{\alpha}_0^i,$$

$$d_n^- = \frac{2}{p^+}\left(\sum_{m=-\infty}^{\infty} d^i_{n-m}\alpha^i_m + \sum_{m=-\infty}^{\infty} S^a_{n-m}\beta^a_m - \frac{1}{6\sqrt{k}} i f_{abc}\sum_{l,m=-\infty}^{\infty} S^a_{n-m-l}S^b_l S^c_m\right)$$

$$\equiv \frac{2}{p^+}F_n, \tag{4.3b}$$

$$\tilde{d}_n^- = \frac{2}{p^+}\tilde{F}_n,$$

where $\tilde{L}_n$ and $\tilde{F}_n$ are the same as L_n and F_n with all the oscillators replaced by the ones with tilde. We now quantize this system and compute the central extension [7, 8a] in the commutator algebra of L_n and F_n. Demanding the closure of the Lorentz algebra requires a special value of this central extension which in turn determines the critical dimension. The quantization proceeds by imposing the following (anti) commutator relations

$$\left[q^i, p^j\right] = i\,\delta^{ij}, \qquad \left[\alpha^i_m, \alpha^j_n\right] = m\,\delta_{m+n,o}\delta^{ij}, \tag{4.4a}$$

$$\left[\beta^a_m, \beta^b_n\right] = \frac{-i}{\sqrt{k}} f^{abc}\beta^c_{m+n} + n\,\delta_{m+n,o}\delta^{ab}, \tag{4.4b}$$

$$\left\{d^i_m, d^j_n\right\} = \delta^{ij}\delta_{m+n,o}, \tag{4.4c}$$

$$\left\{S^a_m, S^b_n\right\} = \delta^{ab}\delta_{m+n,o}. \tag{4.4d}$$

Similar relations hold for tilde oscillators, and the latter commute with those without tilde. In deriving (4.4b) we have used the result of Witten [9] which gives the commutator $[J^a_{+\mathrm{L}}, J^b_{+\mathrm{L}}]$.

We now consider the commutator algebra of L_n and F_n given in (4.3). First we note that these satisfy the following (anti) Poisson brackets:

$$[L_m, L_n]_{\mathrm{P}} = (m-n)L_{m+n},$$

$$[F_m, L_n]_{\mathrm{P}} = \left(m-\tfrac{1}{2}n\right)F_{m+n},$$

$$\{F_m, F_n\}_{\mathrm{P}} = 2L_{m+n} \tag{4.5}$$

and similar expressions for $\tilde{L}_m$ and $\tilde{F}_m$.

These brackets define a super Virasoro algebra with no central extension. In the quantum case one must take care of operator ordering. With the normalization of the individual terms given as in (4.3) one finds, however, that the (anti) commutator algebra of L_n and F_n does not close. By demanding closure, we find that the terms in L_m and F_m have to be normalized differently than in (4.3). The resulting (quantum) expressions are

$$\begin{aligned} L_n = {} & \tfrac{1}{2}\sum_{m=-\infty}^{\infty} :\alpha^i_{n-m}\alpha^i_m: \\ & + \tfrac{1}{2}\sum_{m=-\infty}^{\infty}\left(m-\tfrac{1}{2}n\right):d^i_{n-m}d^i_m: + \frac{1}{2(1+c_{\mathrm{A}}/2k)}\sum_{m=-\infty}^{\infty}:\beta^a_{n-m}\beta^a_m: \\ & + \tfrac{1}{2}\sum_{m=-\infty}^{\infty}\left(m-\tfrac{1}{2}n\right):S^a_{n-m}S^a_m: + \left(\varepsilon(d-2)+\varepsilon' d_{\mathrm{G}}\right)\delta_{n,o}, \end{aligned} \tag{4.6a}$$

$$\begin{aligned} F_n = {} & \sum_{m=-\infty}^{\infty}:d^i_{n-m}\alpha^i_m: + \frac{1}{\sqrt{1+c_{\mathrm{A}}/2k}} \\ & \times\left[\sum_{m=-\infty}^{\infty}:\beta^a_{n-m}S^a_m: - \frac{i}{6\sqrt{k}}f^{abc}\sum_{l,m=-\infty}^{\infty}:S^a_{n-l-m}S^b_m S^c_l:\right], \end{aligned} \tag{4.6b}$$

where

$$p^- = \frac{4}{p^+}(L_0-\alpha_0) = \frac{4}{p^+}(\tilde{L}_0-\alpha_0),$$

$$f_{acd}f_b{}^{cd} = c_{\mathrm{A}}\delta_{ab},$$

$$\varepsilon = \begin{cases} 0, & \text{antiperiodic b.c. for } \lambda \\ \frac{1}{16}, & \text{periodic b.c. for } \lambda, \end{cases}$$

$$\varepsilon' = \begin{cases} 0, & \text{antiperiodic b.c. for } \chi \\ \frac{1}{16}, & \text{periodic b.c. for } \chi. \end{cases} \tag{4.7}$$

In (4.6a) in order to discuss all the boundary conditions simultaneously we have added a constant term to L_0. For the value of c_A, see eq. (1.3). One can now show that L_n and F_n satisfy the super-Virasoro algebra [16]

$$[L_m, L_n] = (m-n)L_{m+n} + \tfrac{1}{12}cm(m^2-1)\delta_{m+n,o}, \tag{4.8a}$$

$$\{F_m, F_n\} = 2L_{m+n} + \tfrac{1}{3}c\left(m^2 - \tfrac{1}{4}\right)\delta_{m+n,o}, \tag{4.8b}$$

$$[F_m, L_n] = \left(m - \tfrac{1}{2}n\right)F_{m+n}, \tag{4.8c}$$

where c is the central extension which is given by

$$c = \tfrac{3}{2}(d-2) + \frac{d_G}{1 + c_A/2k} + \tfrac{1}{2}d_G. \tag{4.9}$$

We now require Lorentz invariance in M_d. The only non-trivial commutator is $[M^{i-}, M^{j-}] = 0$. We find that this commutator holds provided that $\alpha_0 = \frac{1}{2}$ and $c = 12$. Thus follows the following critical dimension formula:

$$d = 10 - \frac{2}{3}\frac{d_G}{1 + c_A/2k} - \tfrac{1}{3}d_G \qquad \text{(spinning string)}. \tag{4.10}$$

In the case of the purely bosonic string where d_m^i and S_m^a are absent (and thus $F_m = 0$), the only surviving commutator is given by (4.8a) with central extension c which now reads

$$c = d - 2 + \frac{d_G}{1 + c_A/2k} \qquad \text{(bosonic string)}. \tag{4.11}$$

The validity of $[M^{i-}, M^{j-}] = 0$ now requires that $\alpha_0 = 1$ and $c = 24$ which implies the following critical dimension formula:

$$d = 26 - \frac{d_G}{1 + c_A/2k} \qquad \text{(bosonic string)}. \tag{4.12}$$

Finally, we quote the quantum mass formula

$$M^2 = 8(L_0 - \alpha_0) - p^i p^i. \tag{4.13}$$

For the case of semisimple group $G = G_1 \times G_2 \times \cdots \times G_P$ the formulae (4.10) and (4.12) are replaced by

$$d = 10 - \tfrac{2}{3}\sum_{i=1}^{P} \frac{d_G^{(i)}}{1 + c_A^{(i)}/2k^{(i)}} - \tfrac{1}{3}\sum_{i=1}^{P} d_G^{(i)} \qquad \text{(spinning string)} \tag{4.14}$$

$$d = 26 - \sum_{i=1}^{P} \frac{d_G^{(i)}}{1 + c_A^{(i)}/2k^{(i)}} \qquad \text{(bosonic string)}. \tag{4.15}$$

5. Mass spectrum

In this section we study the mass spectrum of a bosonic *closed* string moving on (Minkowski)$_d \times$ G, which will illustrate both gravitational as well as Yang-Mills degrees of freedom. We shall comment on the spinning string case at the end of this section.

The quantum mass operator in (Minkowski)$_d$ of the closed bosonic string is given by (see eq. (4.13))

$$\tfrac{1}{4}\alpha' M^2 = N_d + N_G + 1, \tag{5.1}$$

where

$$N_d = \sum_{i=1}^{d-2} \sum_{n=1}^{\infty} \alpha^i_{-n}\alpha^i_n, \tag{5.2}$$

$$N_G = \frac{1}{2(1 + c_A/2k)} \sum_{a=1}^{d_G} \sum_{n=-\infty}^{\infty} :\beta^a_{-n}\beta^a_n:. \tag{5.3}$$

N_d and N_G are the occupation number operators for the left movers satisfying $[N_d, \alpha^i_n] = -n\alpha^i_n, [N_G, \beta^a_n] = -n\beta^a_n$. Analogously one defines the occupation number operators for the right movers, $\tilde{N}_d$ and $\tilde{N}_G$ with α and β replaced by $\tilde{\alpha}$ and $\tilde{\beta}$, respectively. These number operators must satisfy the following closed string constraint:

$$N_d + N_G = \tilde{N}_d + \tilde{N}_G. \tag{5.4}$$

To identify the mass spectrum we need to study the representations of the Kac-Moody algebra [20] of the operators β^a_n. (For a review see ref. [21].) From now on, we shall specialize to the case of simply laced G. (The more complicated case of non-simply laced algebras has been treated in ref. [21].) Furthermore we restrict ourselves to $k = 1$. As for the ordinary Lie algebras every representation is characterized by a highest weight vector. The basis of the representation space is obtained by successive application of the step operators on it. A weight vector obtained in this way is denoted by [20]

$$l = (\Lambda^I, \kappa, \delta), \quad I = 1, \ldots, \text{rank G}, \tag{5.5}$$

where Λ^I are the components of a vector on the weight lattice of G, κ is the eigenvalue of k (see (4.4b)) and δ is the eigenvalue of the derivation operator d [20] which, in our case, is nothing but $(1 - N_G)$.

We shall consider the basic representation [20] which is characterized by the highest weight $l_0 = (0, 1, 1)$. This weight corresponds to the ground state $|0\rangle$. Given the highest weight l_0, the remaining bases of the representation space are obtained by successive application of the appropriate step operators, $\beta^a_{-n} \ldots \beta^b_{-m}|0\rangle$ with

positive $n,\ldots,m$. Using the infinite discrete Weyl group, $\hat{W}$, Frenkel and Kac [29] have shown that the weight vectors of the states obtained in this way must have the form (proposition (2.1) in ref. [29])

$$\left(\alpha^I, 1, 1-p-\tfrac{1}{2}\alpha^I\alpha^I\right), \qquad p=0,1,2,\ldots, \tag{5.6}$$

with degeneracy $M_p(r)$ given by

$$\sum_{p=0}^{\infty} M_p(r)x^p = \prod_{p=1}^{\infty}(1-x^p)^{-r} = 1+rx+\tfrac{1}{2}r(r+3)x^2+\cdots. \tag{5.7}$$

Here r is the rank of G, and α^I is a vector on the root lattice. From (5.6) it follows that $N_G = p + (\frac{1}{2}\alpha^I\alpha^I)$. Thus, eq. (5.1) implies

$$\tfrac{1}{4}\alpha' M^2 = p + \tfrac{1}{2}\alpha^I\alpha^I - 1. \tag{5.8}$$

To illustrate the details of the spectrum let us consider the case of $M_{24}\times SU(3)$ as an example. The states $\beta^a_{-n_1}\beta^b_{-n_2}\cdots\beta^c_{-n_i}|0\rangle$ are eigenstates of $d=1-N_G$ with eigenvalues $1-(n_1+n_2+\cdots+n_i)$. Hence the states $\beta^a_{-1}|0\rangle$ have $d=0$. These have the weights $(\alpha,1,0)$ and $(0,1,0)$ where α is a root of SU(3). The six weights $(\alpha,1,0)$ have unit multiplicity each, while the multiplicity of $(0,1,0)$ is two. Therefore they form an SU(3) octet.

The next set of states (corresponding to $d=-1$) are $\beta^a_{-1}\beta^b_{-1}|\rangle$ and $\beta^a_{-2}|0\rangle$. By virtue of the commutation relations satisfied by the β's, the state $\beta^a_{-2}|0\rangle$ is the antisymmetric part of $\beta^a_{-1}\beta^b_{-1}|0\rangle$. Thus this state is not independent.

The states $\beta^a_{-1}\beta^b_{-1}|0\rangle$ have the weights $(\alpha,1,-1)$ and $(0,1,-1)$ with multiplicities 2 and 5, respectively. They therefore form the representation $1\oplus 8\oplus 8$ of SU(3). Note that the states 10, $\overline{10}$ and 27 in the product 8×8 are not present because their weights are not of the form given in (5.6). One can show that they have vanishing norm. We have checked that the spectrum thus obtained indeed does coincide with the mass spectrum of a bosonic string moving on a direct product of 24-dimensional Minkowski space with a 2-dimensional torus. As mentioned in the introduction, various aspects of this equivalence have been discussed in refs. [7, 8a, 19, 20].

To apply this construction to the case of a closed bosonic string propagating on $M_d\times G$ we need to include $\tilde{\beta}^a_n$ as well as α^i_n and $\tilde{\alpha}^i_n$ in the operator algebra. Since β and $\tilde{\beta}$ generate two commuting Kac-Moody algebras the spectrum has $G\times G$ symmetry.

The ground state $|0\rangle$ here is a singlet of the Poincaré group of M_d as well as of $G\times G$. It is annihilated by all α^i_n, $\tilde{\alpha}^i_n$, β^a_n and $\tilde{\beta}^a_n$ for which $n\geqslant 1$. Hence its mass is given by $\frac{1}{4}\alpha' M^2 = -1$ and therefore it is a tachyonic state.

The first excited level is massless and consists of the following states: $\alpha^i_{-1}\tilde{\alpha}^j_{-1}|0\rangle$ containing a graviton, a second-rank antisymmetric tensor and a dilaton; the states

TABLE 1
The scalars obtained by the operation of β^a_{-n} and $\tilde{\beta}^a_n$ on $|0\rangle$ for the first five levels for G = SU(3)

$\frac{1}{4}\alpha' M^2$	States	SU(3) × SU(3) representation
−1	$\|0\rangle$	(1,1)
0	$\beta^a_{-1}\tilde{\beta}^b_{-1}\|0\rangle$	(8,8)
+1	$\beta^a_{-1}\beta^b_{-1}\tilde{\beta}^c_{-1}\tilde{\beta}^d_{-1}\|0\rangle$	(1,1) + 2(1,8) + 2(8,1) + 4(8,8)
+2	$\beta^{a_1}_{-1}\beta^{\cdot}_{-1}\beta^{\cdot}_{-1}\tilde{\beta}^{\cdot}_{-1}\tilde{\beta}^{\cdot}_{-1}\tilde{\beta}^{a_6}_{-1}\|0\rangle$	4(1,1) + 6(1,8) + 6(8,1) + 2(1,10) + 2(10,1) + 2(1, $\overline{10}$) + 2($\overline{10}$,1) + 9(8,8) + 3(8,10) + 3(10,8) + 3(8, $\overline{10}$) + 3($\overline{10}$,8) + (10,10) + (10, $\overline{10}$) + ($\overline{10}$,10) + ($\overline{10}$, $\overline{10}$)
+3	$\beta^{a_1}_{-1}\beta^{\cdot}_{-1}\beta^{\cdot}_{-1}\beta^{\cdot}_{-1} \times\tilde{\beta}^{\cdot}_{-1}\tilde{\beta}^{\cdot}_{-1}\tilde{\beta}^{\cdot}_{-1}\tilde{\beta}^{a_9}_{-1}\|0\rangle$	(1,1) + 6(1,8) + 6(8,1) + (1,10) + (10,1) + (1, $\overline{10}$) + ($\overline{10}$,1) + (1,27) + (27,1) + 36(8,8) + 6(8,10) + 6(10,8) + 6(8, $\overline{10}$) + 6($\overline{10}$,8) + 6(8,27) + 6(27,8) + (10,10) + (10, $\overline{10}$) + ($\overline{10}$,10) + ($\overline{10}$, $\overline{10}$) + (10,27) + (27,10) + ($\overline{10}$,27) + (27, $\overline{10}$) + (27,27)

$\alpha^i_{-1}\tilde{\beta}^a_{-1}|0\rangle$ and $\tilde{\alpha}^i_{-1}\beta^a_{-1}|0\rangle$ which have spin one and transform in the adjoint representation of G × G; and finally there are the scalars $\beta^a_{-1}\tilde{\beta}^a_{-1}|0\rangle$ which transform as (adj, adj) of G × G. For G = SU(3) in table 1 we have given the spin-zero states obtained by the operation of β and $\tilde{\beta}$ on $|0\rangle$ for the first five levels.

For the closed spinning string the mass formula reads:

$$\tfrac{1}{4}\alpha' M^2 = N^d_{\mathrm{B}} + N^{\mathrm{G}}_{\mathrm{B}} + N^d_{\mathrm{F}} + N^{\mathrm{G}}_{\mathrm{F}} + \Delta\,,$$

where

$$N^d_{\mathrm{F}} = \tfrac{1}{2}\sum_{n=-\infty}^{\infty} n : d^i_{-n} d^i_n :,$$

$$N^{\mathrm{G}}_{\mathrm{F}} = \tfrac{1}{2}\sum_{n=-\infty}^{\infty} n : S^a_{-n} S^a_n :$$

and N^d_{B} and $N^{\mathrm{G}}_{\mathrm{B}}$ are given as before with the constraint

$$N^d_{\mathrm{B}} + N^{\mathrm{G}}_{\mathrm{B}} + N^d_{\mathrm{F}} + N^{\mathrm{G}}_{\mathrm{F}} = \tilde{N}^d_{\mathrm{B}} + \tilde{N}^{\mathrm{G}}_{\mathrm{B}} + \tilde{N}^d_{\mathrm{F}} + \tilde{N}^{\mathrm{G}}_{\mathrm{F}}\,.$$

Corresponding to periodic (P) and antiperiodic (AP) boundary conditions, there are nine possible sectors. However to simplify the discussion we consider only four of these sectors corresponding to the cases where $\lambda^{(1)}$ and $\lambda^{(2)}$ (idem $\chi^{(1)}$ and $\chi^{(2)}$) obey the same boundary conditions. This will not change our results. On table 2 we give Δ as well as the transformation properties of the ground state and the masses of

TABLE 2
The values of Δ, the transformation properties of the ground state and the mass levels for different sectors, corresponding to periodic (P) or antiperiodic (AP) boundary conditions or the fermions λ^α and χ^a

λ^α	χ^a	Δ	ground state	$\frac{1}{4}\alpha' M^2$
AP	AP	$-\frac{1}{2}$	scalar of M_d and G	$-\frac{1}{2}, 0, \frac{1}{2}, \ldots$
AP	P	$\frac{1}{16}(d_G - 8)$	scalar of M_d and spinor of $SO(d_G)$	$\frac{1}{16}(d_G - 8), \frac{1}{16}d_G, \frac{1}{16}(d_G + 8), \ldots$
P	AP	$\frac{1}{16}(d - 10)$	spinor of M_d scalar of G^d	$\frac{1}{16}(d - 10), \frac{1}{16}(d - 2), \frac{1}{16}(d + 6), \ldots$
P	P	$\frac{1}{16}(d + d_G - 10)$	spinor of M_d and $SO(d_G)$	$\frac{1}{16}(d + d_G - 10), \frac{1}{16}(d + d_G + 6), \frac{1}{16}(d + d_G + 22), \ldots$

the first few levels. We can infer from this table that if we desire to have a non-abelian group G we cannot have massless fermions, in agreement with the result of Friedan and Shenker [22]. To clarify this, note that physical fermions must have periodic boundary conditions in Minkowski space-time and therefore are contained in the last two rows of table 2. Thus, supersymmetry in space-time (as contrasted to supersymmetry in $d = 2$) is unlikely to arise in such theories. The only possibility for massless fermions is when $d + d_G = 10$, which when combined with eq. (1.2) gives the unique solution $c_A = 0$. This would permit only a product of U(1)'s (e.g. $U(1)^4$ for $d = 6$)*.

6. Open problems

The results of this paper can be extended in the following directions:

(i) Generalization of our model to spaces other than group manifolds.

(ii) Generalization to $N = 2, 4, 8, 16$ supersymmetries in $d = 2$. The extension of our $N = 1$ model to $N = 2$ has been recently obtained [23], though the critical dimensional formula has not been derived. (See however ref. [24].)

(iii) To study, in any one of the models mentioned above, the possibility of introducing consistent interactions.

In this paper, in d-dimensional space-time we have lost supersymmetry. To maintain it, one may have to consider models of the type constructed by Green and Schwarz [18]. Such models can exist in 3, 4, 6 and 10 dimensions. Recently, Witten has constructed the $d = 10$ Green-Schwarz type action in curved space [25]. Follow-

* Note that there are two possible G-parity [15] operators in the theories of this type, $(-1)^{2N_F^G - 1}$ or $(-1)^{2N_F^G - 1}$. These may be useful for suppressing the tachyons.

ing Witten's method, we have constructed similar actions in $d = 3, 4, 6$ [26]. In this context it is worth remarking that an anomaly-free $N = 2$, $d = 6$ supergravity with $E_6 \times E_7 \times U(1)$ symmetry exists [17] and may provide a good phenomenological model for quarks and leptons. Whether such a theory is a field theoretical limit of a string theory remains to be seen.

We appreciate discussions with L. Alvarez-Gaumé, M. Green, D. Olive, A. Neveu, A. Schwimmer, J. Strathdee and E. Witten.

References

[1] S. Deser and B. Zumino, Phys. Lett. 65B (1976) 369
[2] L. Brink, P. di Vecchia and P. Howe, Phys. Lett. 65B (1976) 471
[3] A. Neveu and J.H. Schwarz, Nucl. Phys. B31 (1971) 86; Phys. Rev. D4 (1971) 1109;
P. Ramond, Phys. Rev. D3 (1971) 2415
[4] J. Wess and B. Zumino, Phys. Lett. 37B (1971) 95;
E. Witten, Nucl. Phys. B223 (1983) 422, 433
[5] S.J. Gates, Jr., C.M. Hull and M. Roček, Nucl. Phys. B248 (1984) 157
[6] T. Curtright and C. Zachos, Phys. Rev. Lett. 53 (1984) 1799;
R. Rohm, Anomalous interactions for the supersymmetric nonlinear σ-model in two dimensions, Princeton preprint (October 1984);
E. Braaten, T.L. Curtright and C.K. Zachos, Nucl. Phys. B260 (1985) 619
[7] D. Nemeschansky and S. Yankielowicz, Critical dimension of string theories in curved space, Phys. Rev. Lett. 54 (1985) 260
[8a] P.G. Freund, Santa Barbara preprint (1984);
V.G. Krizhnik and A.B. Zamolodchikov, Nucl. Phys. B247 (1984) 83;
V.G. Kac, Infinite dimensional Lie algebras, *in* Prog. in Maths. vol. 44 (Birkhäuser, Boston, 1983) p. 245;
P. Goddard, D. Olive and A. Schwimmer, Imperial College preprint TP/84-85/23;
P. Goddard, W. Nahm and D. Olive, Symmetric spaces and chiral field theories in two dimensions, Imperial College preprint (1985)
[8b] S. Jain, R. Shankar and S.R. Wadia, Conformal invariance and string theories in compact space: bosons, Phys. Rev. D32 (1985) 2713
[9] E. Witten, Comm. Math. Phys. 92 (1984) 455
[10] I. Antoniadis and C. Bachas, Conformal invariance and parastatistics in two dimensions, SLAC preprint SLAC-PUB-3625 (April 1985)
[11] F. Ardalan and F. Mansouri, Phys. Rev. D9 (1974) 3341;
J.F.L. Hopkinson and R.W. Tucker, Phys. Rev. D10 (1974) 558
[12] D.J. Gross, J.A. Harvey, E. Martinec and R. Rohm, Nucl. Phys. B256 (1985) 253; B267 (1986) 75
[13] L. Alvarez-Gaumé and D.Z. Freedman, Comm. Math. Phys. 80 (1981) 443
[14] P. Forgács and N.S. Manton, Comm. Math. Phys. 72 (1980) 15
[15] J.H. Schwarz, Phys. Reports 89, (1982) 223
[16] P. Goddard and D. Olive, Nucl. Phys. B257 [FS14] (1985) 226
[17] S. Randjbar-Daemi, A. Salam, E. Sezgin and J. Strathdee, Phys. Lett. 151B (1985) 351;
A. Salam and E. Sezgin, ICTP, Trieste, preprint IC/85/19 (1985)
[18] M.B. Green and J.H. Schwarz, Nucl. Phys. B181 (1981) 502
[19] I. Frenkel and V.G. Kac, Inventiones Math. 62 (1980) 23
[20] D. Olive, Kac-Moody algebras: An introduction for physicists, Imperial College preprint, London (March 1985);

P. Goddard, Infinite-dimensional Lie algebras: representations and applications, Cambridge preprint (1985)
[21] A.J. Feingold and I. Frenkel, Classical affine algebras, preprint
[22] D. Friedan and S. Shenker, Chicago preprint (1985)
[23] E. Bergshoeff, H. Nishino and E. Sezgin, Phys. Lett. 166B (1986) 242.
[24] C.M. Hull, σ-model beta-functions and string compactifications, Cambridge preprint (1985)
[25] E. Witten, Twistor-like transform in ten dimensions, Princeton preprint (May 1985)
[26] E. Bergshoeff, E. Sezgin and P.K. Townsend, Superstring actions in $d = 3, 4, 6, 10$ curved superspace, Trieste preprint (1985), to appear in Phys. Lett. B.

International Journal of Modern Physics A, Vol. 2, No. 3 (1987) 667–693
© World Scientific Publishing Company

σ-MODELS AND STRING THEORIES

S. RANDJBAR-DAEMI*
Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

ABDUS SALAM
Imperial College, London, England and *International Centre for Theoretical Physics, Trieste, Italy*

and

J. A. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 2 March 1987

The propagation of closed bosonic strings interacting with background gravitational and dilaton fields is reviewed. The string is treated as a quantum field theory on a compact 2-dimensional manifold. The question is posed as to how the conditions for the vanishing trace anomaly and the ensuing background field equations may depend on global features of the manifold. It is shown that to the leading order in σ-model perturbation theory the string loop effects do not modify the gravitational and the dilaton field equations. However for the purely bosonic strings new terms involving the modular parameter of the world sheet are induced by quantum effects which can be absorbed into a re-definition of the background fields. We also discuss some aspects of several regularization schemes such as dimensional, Pauli-Villars and the proper-time cut off in an appendix.

Introduction

It is hoped that one or other of the superstring theories[1] will generate a realistic low energy phenomenology.[2] The low energy physics would be described by an effective field theory which must include at least the standard model of electroweak interactions as well as the gravitational interaction. The extraction of such an effective field theory is a technical problem, the general aspects of which are quite well understood.[3] Thus, if the spectrum of string states is known and if the amplitudes involving the scattering of massless states can be constructed, then the low energy behavior of these amplitudes can be reproduced by an effective field theory based on massless fields with specified interactions. By this means it was found, for example, that the gravitational interaction is comprised in certain string theories.

A different way to extract the low energy theory is couched in the language of 2-dimensional conformal field theory or σ-models.[4] In this approach the string amplitudes are not needed. Instead, the putative massless string states are represented by

* Supported by Tomalla-Stiftung und Schweizerischer Nationalfonds

background fields whose arguments are the string coordinates, the dynamical variables of the 2-dimensional world.[5] Out of the coherent amplitudes of the 2-dimensional system it is possible, in principle, to extract the physical amplitudes. For this program to be consistent it is necessary that the 2-dimensional theory should be conformal invariant and this implies restrictions on the background configurations. These restrictions take the form of equations of motion for the background fields and it seems they are governed by an action principle.[4,5] The action for the background fields arrived at in this way defines the low energy effective field theory. It ought to be equivalent to the theory defined by the low energy behavior of the string amplitudes.

Both methods involve lengthy analysis and neither has been carried very far. Their equivalence has yet to be clearly demonstrated. Indeed, there are conceptual uncertainties in the σ-model when it comes to including the effects of string loops. In particular, since the conformal anomaly would seem to be a short-distance, or local effect, one may ask how can it be sensitive to global features of the 2-dimensional space-time? Put another way, a local quantity like the trace anomaly requires a given 2-manifold for its definition. How is one to sum this quantity over various topologies? As far as we are aware, all computations presented up to now have treated the world sheet as a flat infinite space.[6] At a technical level this banishing of global features is very convenient for computing σ-model loop corrections because it permits the use of dimensional regularization. In our view the matter of finding a regularization scheme which both respects the local symmetries (reparametrization invariance) and makes allowance for world sheet topology, is an important unsolved problem. An instance of the kind of difficulty generated by this neglect of topological constraints in σ-model calculations is the infrared singularity that always appears.[7,8,9] This singularity is an artefact which results from replacing the compact sphere with a noncompact plane. We show that the infrared singularity can be dealt with by using heat kernel methods, at least to the orders considered up to now.

The aim of this note is to make a preliminary attempt at σ-model computations on compact manifolds. We shall consider only the bosonic string coupled to background metric tensor and dilaton fields.[5] Using Green's functions appropriate to the surface we shall attempt to calculate the leading contributions to the conformal anomaly. We avoid using the flat space dimensional regularization in this calculation and fall back on a proper-time method[10] to regularize the propagator and define coincidence limits which is probably adequate for the leading order though not for higher orders. Although there is no infrared problem here, there is a normalizable zero-mode of the Laplacian which must be dealt with carefully. This results in the presence of a "gauge" parameter in the Green's functions and it is necessary to show that the conformal anomaly is independent of this parameter. The coincidence limits of the Green's functions do also depend on the modular parameters τ of the surface. Consequently the σ-model vacuum amplitude explicitly contains τ. We show however that the background field equations for the massless modes i.e. the gravity-dilaton system are τ-independent.

Our results for the gravity-dilaton field equations are identical to earlier calculations

based on flat space dimensional regularization.[11] However if a proper-time cut off or Pauli-Villars regularization is used we must define the measure of the path-integral properly in order to agree with the predictions of the dimensional regularization. This is discussed in Sec. 4.

The plan of this paper is as follows. Section 2 contains a brief recapitulation of the σ-model formalism. This includes the leading terms in a normal coordinate expansion[7,8] of the Fradkin-Tsytlin[5] action and a discussion of the Feynman graphs which are relevant. Section 3 gives suitably gauge fixed Green's functions. Here we also consider the various coincidence limits. In Sec. 4 the leading contributions to the conformal anomaly are computed. In Sec. 5 we present our conclusion.

In Appendix A the various regularization schemes are discussed in detail. In Appendix B we give a derivation of the full regularized scalar Green's function on S^2.

2. Generalities

To fix the notation we present here a summary description of the closed bosonic string coupled to background metric tensor and dilaton fields. The dynamical variables are the string coordinates $X^i(\sigma)$, $i = 1, \ldots, D$ and the Euclidean signature metric, $\gamma_{\alpha\beta}(\sigma)$, $\alpha, \beta = 1, 2$ where $\sigma = (\sigma^1, \sigma^2)$ labels a point on the world sheet (which is taken to be compact with Euler number χ). The action functional of Fradkin and Tsetlyn[5,11] is given by

$$S = \frac{1}{4\pi\alpha'} \int_{M_\chi} d^2\sigma \sqrt{\gamma} \left[\gamma^{\alpha\beta} \partial_\alpha X^i \partial_\beta X^j g_{ij}(X) - \frac{\alpha'}{2} R_2(\gamma) \Phi(X) \right]. \tag{2.1}$$

The background fields $g_{ij}(X)$ and $\Phi(X)$ represent the space-time graviton and dilaton, respectively.

In principle one should regard X^i and $\gamma_{\alpha\beta}$ as dynamical variables on the same footing. In practice the metric $\gamma_{\alpha\beta}$ plays a secondary role. Firstly, by exploiting the general covariance it can be reduced, at least locally to the conformally flat form

$$\gamma_{\alpha\beta} = e^{2\rho} \delta_{\alpha\beta} \tag{2.2}$$

where $\rho(\sigma)$ is a real field. The first term of (2.1) is conformally invariant and hence independent of ρ. The second contains ρ explicitly, but this is deceptive. A principal aim of the program is to ensure that ρ drops out of the final amplitudes or, in other words to ensure that the quantized theory is conformal invariant.[4] It is therefore usual to treat ρ as an external field rather than a dynamical one. We shall follow this usage here. (In a suitably regularized theory ρ would probably have to be treated as a truly dynamical variable at least up to the state where renormalized on-shell amplitudes are defined.)

The quantum field theory based on S starts with the geodesic coordinate expansion[7,8] of the string coordinates $X^i(\sigma)$ around some classical configuration $y^i(\sigma)$. One writes

$$X^i(\sigma) = y^i(\sigma) + \sqrt{2\pi\alpha'}\,\xi^i(\sigma) + \cdots \tag{2.3}$$

where $\xi^i(\sigma)$ represents the quantum fluctuations. On substituting the geodesic coordinate expansion (2.3) into (2.1) one obtains

$$S = S_0 + S_2 + S_3 + \cdots \tag{2.4}$$

where S_n includes the terms of order ξ^n,

$$S_0 = \frac{1}{4\pi\alpha'}\int d^2\sigma\sqrt{\gamma}\left[\gamma^{\alpha\beta}\partial_\alpha y^i\partial_\beta y^j g_{ij}(y) - \frac{\alpha'}{2}R_2(\gamma)\Phi(y)\right],$$

$$S_2 = \int d^2\sigma\sqrt{\gamma}\left[\frac{1}{2}\gamma^{\alpha\beta}\nabla_\alpha\xi^a\nabla_\beta\xi^b\delta_{ab} + \frac{1}{2}\gamma^{\alpha\beta}R_{abcd}\nabla_\alpha y^a\xi^b\xi^c\nabla_\beta y^d - \frac{\alpha'}{8}R_2(\gamma)\xi^a\xi^b\nabla_a\nabla_b\Phi\right],$$

$$S_3 = \int d^2\sigma\sqrt{\gamma}\left[\frac{\sqrt{2\pi\alpha'}}{6}R_{abcd;e}\gamma^{\alpha\beta}\nabla_\alpha y^a\nabla_\beta y^d\xi^b\xi^c\xi^e + \frac{2}{3}\sqrt{2\pi\alpha'}R_{abcd}\gamma^{\alpha\beta}\nabla_\alpha y^a\xi^b\xi^c\nabla_\beta\xi^d - \frac{(2\pi\alpha')^{3/2}}{48\pi}R_2(\gamma)\nabla_a\nabla_b\nabla_c\Phi\xi^a\xi^b\xi^c\right],$$

$$S_4 = \int d^2\sigma\sqrt{\gamma}\left[2\pi\alpha'\gamma^{\alpha\beta}\left\{\frac{1}{6}R_{abcd}\nabla_\alpha\xi^a\xi^b\xi^c\nabla_\beta\xi^d + \frac{1}{4}R_{abcd;e}\nabla_\alpha y^a\xi^b\xi^c\nabla_\beta\xi^d\xi^e + \frac{1}{24}(R_{abcd;ef} + 4R_{gbca}R_{gefd})\nabla_\alpha y^a\nabla_\beta y^d\xi^b\xi^c\xi^e\xi^f\right\} - \frac{\pi\alpha'^2}{48}R_2(\gamma)\nabla_a\nabla_b\nabla_c\nabla_d\Phi\xi^a\xi^b\xi^c\xi^d\right], \tag{2.5}$$

etc. To simplify the writing of these expressions we have introduced an orthnormal basis $e_i^a(y)$, $a = 1, 2, \ldots, D$ in the target space. In terms of this basis we have

$$\begin{aligned} g_{ij} &= e_i^a e_j^b \delta_{ab}, \\ \nabla_\alpha y^a &= \partial_\alpha y^i e_i^a, \\ \nabla_\alpha\xi^a &= \partial_\alpha\xi^a + \partial_\alpha y^i\omega_i^{ab}\xi^b, \end{aligned} \tag{2.6}$$

where ω_i^{ab} is the target space (Riemannian) spin connection. Covariant derivatives of target space quantities such as R_{abcd} employ this connection. In (2.5) all dependence on the derivatives $\partial_\alpha y^i$ is explicit.

The first order term, S_1, is not included in (2.4) because, of course, the configuration $y^i(\sigma)$ must minimize the classical action: it must satisfy the equations

$$\frac{1}{\sqrt{\gamma}}\partial_\alpha(\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta y^i) + \gamma^{\alpha\beta}\partial_\alpha y^k \partial_\beta y^l \Gamma^i_{kl}(y) = -\frac{\alpha'}{4}R_2(\gamma)g^{ij}\partial_j\Phi(y) \tag{2.7}$$

where $\Gamma^i_{kl}(y)$ is the target space connection in a coordinate basis. Generally, this equation must be solved on the world sheet M_χ with punctures corresponding to emission vertices. At these punctures suitable boundary data, characterizing the emitted states, would have to be supplied.

Amplitudes in the quantized theory are given by Feynman path integrals,

$$e^{-W} = \text{Const.} \int (\Pi\, d\xi) e^{-S}. \tag{2.8}$$

Thus defined, W is a functional of g_{ij}, Φ and, through $y(\sigma)$, of the boundary data which specifies the scattering process. With respect to the 2-dimensional quantum fields, ξ^i, the functional W is just the connected part of the vacuum amplitude.[10] The first few terms in its loop expansion are shown in Fig. 1.

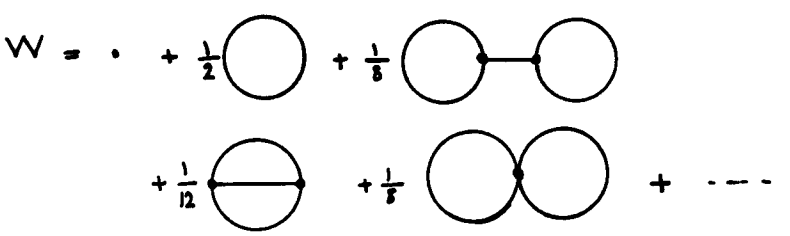


Fig. 1. Graphical representation of the connected part of the vacuum amplitude for the 2-dimensional quantum field theory defined by the action (2.4), (2.5). Lines represent the propagator corresponding to S_2, the single dot represents the classical contribution, S_0, and the circle without vertices represents the one-loop contribution in Det S_2.

The propagators and vertices needed for the construction of W are defined by the functionals, $S_2, S_3, \ldots$ listed above. In particular, the propagator defined by the path integral

$$G^{ab}(\sigma,\sigma') = \int \Pi(d\xi)\xi^a(\sigma)\xi^b(\sigma')e^{-S_2} \tag{2.9a}$$

must satisfy the equation

$$\left[-\delta_{bc}\nabla^2 + \gamma^{\alpha\beta}R_{abcd}\nabla_\alpha y^a \nabla_\beta y^d - \frac{\alpha'}{4}R_2(\gamma)\nabla_b\nabla_c\Phi\right]G^{cf}(\sigma,\sigma') = \frac{1}{\sqrt{\gamma}}\delta_b^f\delta_2(\sigma,\sigma') - \text{const.} \tag{2.9b}$$

where $\delta_2(\sigma,\sigma')$ is the delta function on the manifold M. It is a scalar density.

To justify the loop expansion it is necessary that the successive terms in (2.4) should be small. Presumably this is obtained when the background fields, g_{ij} and Φ, are slowly varying on the scale of the slope parameter, α'. We shall assume that $\sqrt{\alpha'}g_{ij,k}$ and $\sqrt{\alpha'}\partial_k\Phi$ are small. We shall also assume that the derivatives $\partial_\alpha y^i$ are small in the sense that $b\partial_\alpha g_{ij}(x)$ and $b\partial_\alpha\Phi(y)$ are small where b measures the "size" of the world sheet, $R_2(\gamma) \sim 1/b^2$. Of course, this is a somewhat artificial consideration since the final theory is supposed to be conformal invariant, i.e. independent of b, but it gives a useful ordering principle at the intermediate stages of the calculations. With these assumptions it is straightforward to obtain the leading terms in the expansion of W in powers of the small quantities

$$\alpha' R_{abcd}, \quad \sqrt{\alpha'}\partial_i\Phi \quad \text{and} \quad (b\nabla_\alpha y)^2 R_{abcd}.$$

The results are collected in Sec. 4.

A necessary ingredient for these computations is an explicit representation for the Green's function on M_χ. There are some technical subtleties connected with the separation of zero-modes. These are reviewed in Sec. 3.

3. Green's Functions

The Green's function defined by (2.9) involves the background fields, g_{ij} and Φ. If these fields are slowly varying then an expansion is called for. The zeroth order term will be essentially the kernel of the Laplace-Beltrami operator on a compact manifold. The purpose of this section is to review some of its properties and derive its various coincidence limits.

The first point to observe is the that the path integral (2.9a) is ill-defined owing to the presence of zero-modes. The action S_2, in the absence of background fields, is invariant under the shift,

$$\xi^a(\sigma) \to \xi^a(\sigma) + c^a \tag{3.1}$$

where the c^a, $a = 1, \ldots, D$ are independent of σ. This means, for example, that the average over M_χ of $\xi^a(\sigma)$ is not determined. It is a "gauge dependent" quantity. The gauge transformations (3.1) are quite trivial, and one hardly needs the Faddeev-Popov armoury to deal with them. Indeed it is simple to show that a suitable gauge fixed action is given by

$$S_2 = \frac{1}{2}\int d^2\sigma\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\alpha\xi^a\partial_\beta\xi^a + \frac{\beta}{2}\left[\int d^2\sigma\sqrt{\gamma}\xi^a\right]^2 \tag{3.2}$$

where β is a constant. One must ensure that the gauge parameter, β, does not appear in any observable quantities.[a]

The Green's function corresponding to (3.2) can be represented by the path integral

$$G^{ab}(\sigma,\sigma') = \int (\Pi d\xi)\xi^a(\sigma)\xi^b(\sigma')e^{-S_2}. \tag{3.3}$$

A more precise expression is obtained by expanding the fields in a discrete orthonormal basis,

$$\xi^a(\sigma) = \sum_n \xi_n^a u_n(\sigma) \tag{3.4a}$$

where the coefficients ξ_n^a are defined by

$$\xi_n^a = \int_{M_\chi} d^2\sigma\sqrt{\gamma}u_n(\sigma)\xi^a(\sigma), \quad n = 0, 1, \ldots. \tag{3.4b}$$

Choose the functions $u_n(\sigma)$ to diagonalize the action (3.2), i.e.

$$-\frac{1}{\sqrt{\gamma}}\partial_\alpha(\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta u_n) = \lambda_n^2 u_n.$$

The zero mode, $n = 0$, is given by

$$u_0(\sigma) = \frac{1}{\sqrt{V}}$$

where V is the area of the manifold, M_χ. The action takes the form,

$$S_2 = \frac{1}{2}\sum_{n>0}\lambda_n^2(\xi_n^a)^2 + \frac{\beta V}{2}(\xi_0^a)^2.$$

Integration of (3.3) then leads directly to the series representation,

$$G^{ab}(\sigma,\sigma') = \delta^{ab}\left(\sum_{n>0}\frac{1}{\lambda_n^2}u_n(\sigma)u_n(\sigma') + \frac{1}{\beta V}\right) \tag{3.5}$$

in which the gauge dependence of G is manifest. It is clear that this Green's function

[a] Although it will not be important for the limited calculations of Sec. 4, we should point out that the transformation (3.1) is a symmetry of the free action only. It is the linearized remnant of the target space general coordinate transformations with respect to which the action (2.1) is invariant. The zero modes would in general correspond to symmetries of the background field configuration (killing vectors of $g_{ij}(X)$, $\Phi(x)$).

is symmetric under the interchange of σ with σ' and that it satisfies

$$-\frac{1}{\sqrt{\gamma}}\partial_\alpha(\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta G) = \sum_{n>0} u_n(\sigma)u_n(\sigma')$$

$$= \frac{1}{\sqrt{\gamma}}\delta_2(\sigma,\sigma') - \frac{1}{V}. \tag{3.6}$$

Now consider the case of the sphere. Here it is quite useful to project onto the complex plane, defining

$$z = \sqrt{2}b\cot\frac{\theta}{2}e^{i\varphi}$$

where b denotes the radius of the sphere and θ, φ are the usual polar angles. With this parametrization the south pole, $\theta = \pi$, maps to the origin, $z = 0$, and the north pole, $\theta = 0$, maps to the point at infinity. The line element takes the form

$$ds^2 = \frac{2dz\,d\bar{z}}{\left(1+\frac{|z|^2}{2b^2}\right)^2}. \tag{3.7}$$

Equation (3.6) now takes the form

$$-2\partial_z\partial_{\bar{z}}G(z,\bar{z}) = \delta_2(z,\bar{z}) - \frac{1}{4\pi b^2\left(1+\frac{|z|^2}{2b^2}\right)^2}. \tag{3.8}$$

It is solved by[b]

$$G(z,z') = -\frac{1}{4\pi}\ln\frac{|z-z'|^2}{\left(1+\frac{|z|^2}{2b^2}\right)\left(1+\frac{|z'|^2}{2b^2}\right)} + \text{const.} \quad z \neq z' \tag{3.9a}$$

$$G(z,z)|_{z=z'} = -\frac{1}{4\pi}\ln\varepsilon, \tag{3.9b}$$

where ε is a short distance cut off. (The form (3.9b) is dictated by the rotational symmetry of S^2.)

As was to be expected this propagator is bounded for $z \to \infty$, but has the usual short distance behavior. The constant could be evaluated by integrating (3.9) over the sphere and comparing with (3.5).

[b] See Appendix B for derivations.

Consider the case of the torus. The torus can be represented by a parallelogram in the complex plane

$$z = \sigma^1 + \tau\sigma^2 \quad \mathrm{Im}\,\tau > 0, 0 < \sigma^1, \sigma^2 \leq 1$$

with opposite sides identified. The line element is flat,

$$ds^2 = dz\, d\bar{z}. \tag{3.10}$$

The area of the torus is given by

$$V = \int_0^1 d\sigma' \int_0^1 d\sigma^2 \sqrt{\gamma}$$

$$= \tau_2. \tag{3.11}$$

Equation (3.6) for the Green's function takes the form

$$-4\partial_z \partial_{\bar{z}} G(z, z') = \frac{1}{\tau_2}\delta_2(z, z') - \frac{1}{\tau_2}. \tag{3.12}$$

It must be solved subject to periodicity conditions, $z \to z + 1$ and $z \to z + \tau$. In terms of real coordinates (σ^1, σ^2), the solution can be represented as a double Fourier series,

$$G(z, z') = \frac{\tau_2}{(2\pi)^2} \sum{}' |\tau n_1 - n_2|^{-2} \exp 2\pi i(n_1\sigma^1 + n_2\sigma^2) + \text{const}. \tag{3.13}$$

where the sum over integers n_1, n_2, excludes the point $n_1 = n_2 = 0$. This result can be expressed in a manifestly modular invariant[c] form,

$$G(z, z') = \frac{1}{2\tau_2}(\mathrm{Im}(z - z'))^2 - \frac{1}{2\pi} \ln \left| \frac{\theta_1(z - z', \tau)}{\eta(\tau)} \right| \quad z \neq z'$$

$$G(z, z, \varepsilon) = -\frac{1}{4\pi} \ln \varepsilon - \frac{1}{\pi} \ln |\eta(\tau)|, \tag{3.14}$$

[c] The modular group of the torus is generated by two elements with respect to which τ and z transform as $(\tau, z) \to (\tau + 1, z)$ and $(-1/\tau, z/\tau)$. The invariance of (3.14) follows from the identities,[12]

$$\theta_1(z, \tau + 1) = \exp(i\pi/4)\theta_1(z, \tau); \quad \eta(\tau + 1) = \eta(\tau)$$

$$\theta_1(z/\tau, -1/\tau) = -(-i\tau)^{1/2} \exp(i\pi z^2/\tau)\theta_1(z, \tau)$$

$$|\eta(-1/\tau)| = |\tau^{1/2}\eta(\tau)|$$

where θ_1 is the odd Jacobi theta function and $\eta(\tau)$ is the Dedkind function, defined by

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n)$$

where $q = \exp(2\pi i\tau)$.

For the lowest order calculation of Sec. 4, the full propagator is not needed, only some coincidence limits. These require the adaptation of some regularization scheme. Since most of the specific aspects of the string physics are intimately related to the 2-dimensionality of the world sheet it is preferable to use a scheme of regularization which preserves this basic feature. The heat kernel method is ideally suited for this, as the compactness of the surface and the presence of the normalizable zero modes can be readily treated in this scheme. (Other regularizations are discussed in the Appendices A and B.)

Following Polyakov[15] we define the regularized propagator by

$$G(\sigma, \sigma'; \varepsilon) = \sum_{n \neq 0} \frac{u_n(\sigma) u_n(\sigma')}{\lambda_n^2} e^{-\varepsilon \lambda_n^2}. \tag{3.16a}$$

In the conformal coordinates (2.2) the ρ-dependence of $G(\sigma, \sigma; \varepsilon)$ has been worked out by Polyakov.[15] From the calculations on S^2 and T^2 it seems likely that on any surface (see also Appendix A)

$$G(\sigma, \sigma; \varepsilon) = -\frac{1}{4\pi} \ln \varepsilon + \frac{1}{2\pi} \rho(\sigma) + g(\tau) \tag{3.16b}$$

where $g(\tau)$ is independent of σ and ε and depends only on the modular parameters. For instance comparison with (3.14) indicates that for T^2

$$g(\tau) = -\frac{1}{\pi} \ln |\eta(\tau)|.$$

For a higher genus surface $g(\tau)$ can be read from the explicit formulae of Ref. 17.

Next we show that $G(\sigma, \sigma; \varepsilon)$ and the coincidence limit of the heat kernel are the only ingredients needed to obtain the derivatives of G in the coincidence limit.

From (3.16a) it follows that

$$\nabla_\alpha G(\sigma, \sigma'; \varepsilon)|_{\sigma=\sigma'} = \frac{1}{2} \nabla_\alpha G(\sigma, \sigma, \varepsilon) \tag{3.17a}$$

$$\nabla_\alpha \nabla'_\alpha G(\sigma, \sigma'; \varepsilon)|_{\sigma=\sigma'} = \frac{1}{2} \nabla^2 G(\sigma, \sigma, \varepsilon) + K'(\sigma, \sigma; \varepsilon) \tag{3.17b}$$

where $K'(\sigma,\sigma';\varepsilon)$ is defined by

$$K'(\sigma,\sigma';\varepsilon) = \sum_{n\neq 0} u_n(\sigma)u_n(\sigma')e^{-\varepsilon\lambda_n^2}$$

$$= K(\sigma,\sigma';\varepsilon) - u_0^2. \tag{3.18a}$$

As $\sigma \to \sigma'$ and $\varepsilon \to 0$ we have[10]

$$K(\sigma,\sigma';\varepsilon) = \frac{1}{4\pi\varepsilon} e^{-d(\sigma,\sigma')/4\varepsilon} \sum_{l=0}^{\infty} \varepsilon^l H_l(\sigma,\sigma') \tag{3.18b}$$

where $d(\sigma,\sigma')$ is the geodesic distance between σ and σ' and $H_n(\sigma,\sigma')$ are regular functions with well-defined coincidence limits. In 2-dimensions we only need $H_0(\sigma,\sigma) = 1$ and $H_1(\sigma,\sigma) = \frac{1}{6}R_2(\gamma)$.

The constant u_0^2 in (3.18a) is given by the volume of the surface, i.e. $u_0^2 = \frac{1}{V}$. As we saw in (3.11) for T^2, $V = \tau_2$. Thus the relevant part of (3.18b) is given by

$$K(\sigma,\sigma;\varepsilon) = \frac{1}{4\pi\varepsilon} + \frac{1}{24\pi}R_2(\gamma) - \frac{1}{V} + O(\varepsilon). \tag{3.19}$$

Now we summarize the relevant coincidence limits

$$G(\sigma,\sigma;\varepsilon) = -\frac{1}{4\pi}\ln\varepsilon + \frac{1}{2\pi}\rho(\sigma) + g(\tau), \tag{3.20a}$$

$$\partial_\alpha G(\sigma,\sigma';\varepsilon)|_{\sigma=\sigma'} = \frac{1}{4\pi}\partial_\alpha\rho, \tag{3.20b}$$

$$\partial_\alpha\partial'_\alpha G(\sigma,\sigma';\varepsilon)|_{\sigma=\sigma'} = \frac{1}{4\pi}\partial^2\rho + e^{2\rho}K'(\sigma,\sigma;\varepsilon) \tag{3.20c}$$

$$= \frac{1}{6\pi}\partial^2\rho + \frac{e^{2\rho}}{4\pi\varepsilon} - \frac{e^{2\rho}}{V}. \tag{3.20d}$$

The coefficient of $\partial^2\rho$ on the right-hand side of (3.20d) coincides with a derivation based on Pauli-Villars regularization. It disagrees with the one obtained from the application of the dimensional regularization (see Ref. 13 and the appendices).

The leading terms in the vacuum functional W, to be described in Sec. 4, will receive τ-dependent contributions from (3.20). However we shall show that τ does not appear in the graviton-dilaton field equation.

4. Trace Anomaly

Our purpose here is to examine the leading contributions to the connected vacuum amplitude W, defined in Sec. 2. The lowest order graphs are shown in Fig. 1. The zeroth-order term, indicated by a dot in Fig. 1, is just the classical contribution, S_0, defined in (2.5). The one-loop graph represents $\ln \operatorname{Det} \Delta$ and this can be expanded in powers of the background R and $\nabla^2\Phi$. Terms linear in these quantities are indicated in Figs. 2a and 2b, respectively. Another term, linear in R, is indicated by the 2-loop graph in Fig. 2c. A contribution to the anomaly which is independent of the background fields also arises here—this is to be cancelled by the contribution of reparametrization ghosts[15] when $D = 26$ and will be ignored in the following.

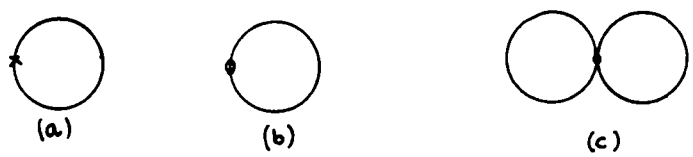


Fig. 2. First order contributions to the vacuum amplitude. The Green's function $G^{ab}(\sigma, \sigma')|_{\omega=0}$ is represented by a solid line. The vertices in (a) and (b) indicate insertions of the operators $\frac{1}{2}R_{abcd}\nabla_e y^a \xi^b \xi^c \nabla_e y^d$ and $\frac{\alpha'}{4}\partial^2 \rho \xi^b \xi^c \nabla_b \nabla_c \Phi$, respectively. The vertex in (c) corresponds to $\pi\frac{\alpha'}{3}R_{abcd}\nabla_e \xi^a \xi^b \xi^c \nabla_e \xi^d$.

Our main objective is to evaluate the trace anomaly $\dfrac{\delta W}{\delta \rho(\sigma)}$. In general this quantity should receive contributions from both the variation of the measure as well as the variation of the ρ-dependence of the graphs mentioned above. The latter originates from the dependence of $G(\sigma, \sigma; \varepsilon)$ on $\rho(\sigma)$ and is the only source of the trace anomaly in dimensional regularization. However if we use a proper-time cut off the changes of the integration measure under $\rho(\sigma) \to \rho(\sigma) + \delta\rho(\sigma)$ must also be accounted for. To this end we must reconsider the formal expression $\prod_{\sigma,a} d\xi^a(\sigma)$ in (2.8) which is assumed to have descended from $\prod_{\sigma,i} dX^i(\sigma)$.

Before the expansion (2.3) the measure must be invariant under the general coordinate transformations of the target space. Hence for a fixed σ we must consider the invariant volume element $d^D X(\sigma)\,(\operatorname{Det} g_{ij}(X(\sigma)))^{1/2}$. Now by expanding $g_{ij}(X(\sigma))$ in powers of (2.3), viz.

$$g_{ij}(X(\sigma)) = g_{ij}(y(\sigma)) + \frac{2\pi\alpha'}{3}R_{iklj}\xi^k(\sigma)\xi^l(\sigma) + \cdots$$

we see that up to the relevant order in α' we can write

$$\operatorname{Det} g_{ij}(X(\sigma)) = \exp\frac{-2\pi\alpha'}{3}R_{ab}\xi^a(\sigma)\xi^b(\sigma). \tag{4.1}$$

Next we consider[18] the variation of $\prod_{n,a} d\xi_n^a$. Under $\rho(\sigma) \to \rho(\sigma) + \delta\rho(\sigma)$ the Laplacian $\Delta = -e^{-2\rho}\partial^2$ changes to $\Delta - 2\delta\rho\Delta$ and hence its eigenfunctions $u_n(\sigma)$ change to $u_n(\sigma) + \delta u_n(\sigma)$. Since $\xi^a(\sigma)$ is neutral, it follows that

$$\xi^a(\sigma) = \sum_n \xi_n^a u_n(\sigma)$$

$$= \sum_n \xi'^a_n \big(u_n(\sigma) + \delta u_n(\sigma)\big).$$

Using the orthonormality of the eigenfunctions u_n we can evaluate the Jacobian determinant and obtain

$$\mathrm{Det}\left(\frac{\partial \xi_n^a}{\partial \xi'^b_m}\right) = \mathrm{Det}\left[\delta_b^a\left(\delta_{mn} - \int d^2\sigma e^{2\rho}\delta\rho(\sigma)u_n(\sigma)u_m(\sigma)\right)\right]$$

$$= \exp\left[-\int d^2\sigma\sqrt{\gamma}\delta\rho(\sigma)\sum_{n\neq 0} u_n(\sigma)u_n(\sigma)\right].$$

The sum in the exponent can be defined by

$$\sum_{n\neq 0} u_n(\sigma)u_n(\sigma) = \lim_{\varepsilon\to 0}\sum_{n\neq 0} u_n(\sigma)u_n(\sigma)e^{-\varepsilon\lambda_n^2}$$

$$= \lim_{\varepsilon\to 0} K'(\sigma,\sigma;\varepsilon)$$

where $K'(\sigma,\sigma;\varepsilon)$ is the heat kernel. Hence under $\rho(\sigma) \to \rho(\sigma) + \delta\rho(\sigma)$

$$\prod_{n,a} d\xi_n^a \to \prod_{n,a} d\xi_n^a e^{-\int d^2\sigma\sqrt{\gamma}\delta\rho(\sigma)K'(\sigma,\sigma;\varepsilon)}. \tag{4.2}$$

Up to the relevant order in α' the combined effect of (4.1) and (4.2) can be summarized by the following prescription:

(i) Replace the action S in (2.8) by S'

$$S' = S - a\frac{\pi\alpha'}{3}\int d^2\sigma\sqrt{\gamma}K'(\sigma,\sigma;\varepsilon)R_{ab}\xi^a(\sigma)\xi^b(\sigma), \tag{4.3}$$

where $a = 0$ in dimensional regularization and $a = 1$ for a proper time cut off and Pauli-Villars scheme. In (4.3) S is given as before by (2.4) and (2.5).

(ii) In evaluating $\dfrac{\delta W}{\delta\rho}$ ignore the changes of $\prod_{n,a} d\xi_n^a$.

With these preliminaries we can write the regularized expression for W.

$$W = W_0(\tau) + \int d^2\sigma \Big[\frac{1}{4\pi\alpha'} \Big(\sqrt{\gamma}\gamma^{\alpha\beta}\nabla_\alpha y^a \nabla_\beta y^a - \frac{\alpha'}{2}\sqrt{\gamma}R_2(\gamma)\Phi(y) \Big)$$

$$+ \frac{2\pi\alpha'}{3}\sqrt{\gamma}K'(\sigma,\sigma;\varepsilon)R_{ab}G^{ab}(\sigma,\sigma;\varepsilon)$$

$$+ \frac{1}{2}\sqrt{\gamma}\gamma^{\alpha\beta}\nabla_\alpha y^a \nabla_\beta y^d R_{abcd}G^{bc}(\sigma,\sigma;\varepsilon)$$

$$- \frac{\alpha'}{8}\sqrt{\gamma}R_2(\gamma)\nabla_b\nabla_c\Phi G^{bc}(\sigma,\sigma;\varepsilon)$$

$$+ \frac{\pi\alpha'}{3}\sqrt{\gamma}\gamma^{\alpha\beta}R_{abcd}\big(\partial_\alpha\partial'_\beta G^{ad}(\sigma,\sigma';\varepsilon)G^{bc}(\sigma,\sigma';\varepsilon)$$

$$+ \partial_\alpha G^{ac}(\sigma,\sigma';\varepsilon)\partial_\beta G^{db}(\sigma,\sigma';\varepsilon)\big)_{\sigma=\sigma'} \Big]$$

$$= W_0(\tau) + \int d^2\sigma \Big[\frac{1}{4\pi\alpha'}(\partial_\alpha y^i \partial_\alpha y^j g_{ij} + \alpha'\partial^2\rho\Phi)$$

$$- \frac{1}{2}\partial_\alpha y^i \partial_\alpha y^j R_{ij}G(\sigma,\sigma;\varepsilon) + \frac{\alpha'}{4}\partial^2\rho\nabla^2\Phi G(\sigma,\sigma;\varepsilon)$$

$$+ \frac{\pi\alpha'}{3}R\Big(-G_{\alpha\alpha'}(\sigma,\sigma';\varepsilon)G(\sigma,\sigma;\varepsilon) + G_\alpha(\sigma,\sigma';\varepsilon)G_\alpha(\sigma,\sigma;\varepsilon)$$

$$+ \sqrt{\gamma}K'(\sigma,\sigma;\varepsilon)G(\sigma,\sigma;\varepsilon)\Big)_{\sigma=\sigma'} \Big], \tag{4.4}$$

where W_0 is the partition function for a flat target space and vanishing Φ. In this formula we have restored the target space coordinates y^i and metric g_{ij}. The Ricci tensor is defined by $R_{ij} = R_{ikj}{}^k$.

The coincidence limits appearing here are to be read from (3.20). The result is

$$W = W_0 + \int d^2\sigma \Big[\frac{1}{4\pi\alpha'}(\partial_\alpha y^i \partial_\alpha y^j g_{ij} + \alpha'\partial^2\rho\Phi(y))$$

$$+ \Big(-\frac{1}{2}\partial_\alpha y^i \partial_\alpha y^j R_{ij} + \frac{\alpha'}{4}\partial^2\rho\nabla^2\Phi \Big)\Big(-\frac{1}{4\pi}\ln\varepsilon + \frac{1}{2\pi}\rho + g(\tau) \Big)$$

$$+ \frac{\alpha'}{16\pi}R\partial_\alpha\rho\partial_\alpha\rho \Big]. \tag{4.5}$$

The ε-dependent terms can be absorbed in a renormalization of the background fields. However up to the relevant order in α' the conditions for conformal invariance, $\dfrac{\delta W}{\delta \rho} = 0$ are insensitive to these renormalizations.[d]

To evaluate $\dfrac{\delta W}{\delta \rho}$ it is necessary to bear in mind that the ρ-dependence in (4.5) is not entirely explicit. The classical configuration $y^i(\sigma)$ depends explicitly on ρ because it must solve the equation (2.7) or

$$\partial^2 y^i + \partial_\alpha y^k \partial_\alpha y^l \Gamma^i_{kl} = \frac{\alpha'}{2} \partial^2 \rho g^{ij} \partial_j \Phi \tag{4.6}$$

in which ρ appears explicitly albeit weakly. To the lowest order the explicit ρ-dependence can be ignored and the variational derivative of W takes the form

$$\frac{\delta W}{\delta \rho} = \frac{1}{4\pi}(\partial^2_\alpha \Phi - \partial_\alpha y^i \partial_\alpha y^j R_{ij}) + \frac{\alpha'}{4\pi}\left(-\frac{1}{2}R + \nabla^2 \Phi\right)\partial^2 \rho. \tag{4.7}$$

It remains to use (4.6) to bring out the ρ-dependence of $\partial^2_\alpha \Phi$

$$\partial^2_\alpha \Phi = \nabla_i \nabla_j \Phi \partial_\alpha y^i \partial_\alpha y^j + \frac{\alpha'}{2} \partial^2 \rho (\nabla \Phi)^2 \tag{4.8}$$

substitute in (4.7) and set $\dfrac{\delta W}{\delta \rho} = 0$ to obtain

$$R_{ij} - \nabla_i \nabla_j \Phi = 0 \tag{4.9a}$$

$$\tfrac{1}{2}R - \nabla^2 \Phi - \tfrac{1}{2}(\nabla \Phi)^2 = 0. \tag{4.9b}$$

[d] In (4.5) we can absorb all τ-dependence in the renormalization counter-term, viz.

$$g^R_{ij} = g_{ij} + \frac{\alpha'}{2} R_{ij}(\ln \varepsilon - 4\pi g(\tau))$$

$$\Phi^R = \Phi(y) - \frac{\alpha'}{4} \nabla^2 \Phi (\ln \varepsilon - 4\pi g(\tau))$$

Up to the relevant order in α' the background equations for g^R_{ij} and Φ^R derived from $\dfrac{\delta W^R}{\delta \rho} = 0$ are identical to (4.9). The above counterterms as well as W_0 in (4.5) will yield a new infinity due to the τ-integration. It has been suggested in Ref. 19 that the regularization of this infinity may be one possible way of incorporating the string loop effects in the low energy field theory.

Thus—up to the leading order in the σ-model perturbation theory—the graviton-dilaton field equations are independent of τ and hence they will not reflect the effects of the string loop corrections.

5. Discussion

Our purpose in this work has been to study some of the features of σ-model calculations on compactified world sheets. If global properties of the world sheet are significant for the low energy theory then it is essential to have regulators that are compatible with modular invariance. In general the σ-model vacuum amplitude, W, depends explicitly on the moduli of the world sheet. In the lowest σ-model order, however, this dependence drops out of the part of the conformal anomaly containing the massless modes of the string. Hence if the vanishing of the trace anomaly is to yield the low energy field equations for the massless states of the string these equations—at least up to the order considered—will be insensitive to the string loop corrections.

Within the closed bosonic strings the τ-integral of W over the modular space is infinite. It is known that for a flat target space the divergences are related to the presence of the tachyon as well as the dilaton in the spectrum. In a curved target space the τ-dependence of W can be absorbed in the renormalization counterterms of both Φ and g_{ij}. The τ-integral of the infinite counterterms will therefore interrelate the σ-model divergences with the string loop infinities.

In the string application of the σ-model the ε-dependent part of W can also be removed by a trivial rescaling of $\gamma_{\alpha\beta}$, namely by replacing ρ by $\rho + \frac{1}{2}\ln\varepsilon$. This amounts to a simple change of the integration variable.[5] It is not difficult to check that the $\dfrac{\delta W}{\delta \rho}$ is unaltered and hence we obtain the same gravity-dilation field equations as before.

The possibility of eliminating the ε-dependent terms from W indicates that the only sources of divergences in the string scattering amplitudes are the modular integrals. The σ-model infinities are just the artefact of a two-step process of calculation. We expect that the above prescription, applied to a superstring theory, should yield a completely unambiguous and finite expression for the generating functional of the S-matrix elements defined by

$$\Gamma = \sum_L \int_F d\mu(\tau) \int (d\rho) e^{-W(\rho,\tau)} \tag{5.1}$$

where F denotes the fundamental domain in the space of the moduli.

It may turn out that the most efficient way to extract the low energy physics is by examining the low energy behavior of amplitudes for massless string states in a flat background.[3] Generally there are two free parameters, the slope, α', and the loop counting parameter, g. A typical bosonic amplitude in ten dimensions, with E external lines would take the form of a sum over loop contributions,

$$T(k_1,\ldots,k_E)=\sum_L g^{E+2L-2}\alpha'^{2E-5}F_L(\sqrt{\alpha'}k_1,\ldots). \tag{5.2}$$

The dependence on α' is governed by dimensional considerations.[e] To single out the gravitational part, if it can be shown that these amplitudes are consistent with general covariance then their low energy forms must be obtainable from an equivalent generally covariant field theory. Schematically, this theory would be characterized by the action functional,

$$I\sim\int d^{10}x\left[\frac{1}{\alpha'^4g^2}R(a_0+a_1g^2+a_2g^4+\cdots)+\frac{1}{\alpha'^3g^2}R^2(b_0+b_1g^2+\cdots)\right.$$
$$\left.+\frac{1}{\alpha'^2g^2}R^3(c_0+c_1g^2+\cdots)+\cdots\right]. \tag{5.3}$$

The numerical coefficients, a_L, b_L, ... can in principle be obtained by comparing amplitudes computed on the basis of (5.3) with the L loop string amplitudes (5.2). Of course these amplitudes are not yet well understood for $L\geq 2$, and the utility of this approach remains to be seen.

The approach to background field equations by examining the ultraviolet infinities of σ-models, up to now, has been formulated on a flat noncompact world sheet.[6] If the program is extended to compact manifolds one expects to find new infinities associated with the integration over modular variables. The connection, if any, of these infinities to the background consistency problem remains to be clarified. In any event, to our knowledge, no proposal for obtaining the higher string-loop coefficients,[19] a_L, b_L, ..., through an examination of the ultraviolet properties of σ-models has yet appeared.

Perhaps one of the more inconvenient aspects of the approach discussed in Sec. 2 is the involvement of the classical configuration $y^i(\sigma)$ at every step. This is of course a well known feature of the background field method. It could be avoided if the variables $\rho(\sigma)$ and $X^i(\sigma)$ were treated on the same footing. One could then define propagators for both ρ and X^i by expanding the classical action $S(\rho,X)$ around an *arbitrary* point in configuration space. The quantum action,[10] $\Gamma(\rho,X)=S+\frac{\hbar}{i}\ln\operatorname{Det}G+\cdots$, would then correspond to the sum of *irreducible* vacuum graphs. The advantage of this program is that the condition for conformal invariance—including, perhaps, ultraviolet finiteness—would be expressed by the equations, $\frac{\delta\Gamma}{\delta\rho(\sigma)}=0$, which would have to be identities in $X^i(\sigma)$ (at least for world sheets without punctures). Before such a program can be formulated, however, it will be necessary to have a suitably regularized

[e] The canonical dimension of a $D=10$ bosonic field is 4. The dimensions of g and α' are 0 and -2, respectively.

$S(\rho, X)$ so that $\rho(\sigma)$ will not simply decouple as it does from the unregularized classical action.

Acknowledgment

At various stages of this work we have benefited from useful discussion with S. de Alwis, J. Govaerts, H. Leutyler and A. Namazie.

Appendix A On Regularization Schemes

In this appendix we calculate regularized expressions for $G(\sigma,\sigma)$, $\nabla_\alpha G(\sigma,\sigma')|_{\sigma=\sigma'}$ and $\gamma^{\alpha\beta}\nabla_\alpha\nabla'_\beta G(\sigma,\sigma')|_{\sigma=\sigma'}$. We shall show that for $\nabla\nabla' G(\sigma,\sigma')|_{\sigma=\sigma'}$ the heat kernel methods yield results in agreement with Pauli-Villars regularization and disagreement with the predictions of the dimensional regularization.

A.1. *The basic equations*

On a compact 2-dimensional manifold the Laplace-Beltrami operator is defined by

$$\Delta = -\frac{1}{\sqrt{\gamma}}\partial_\alpha(\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta) \tag{A.1a}$$

with the following eigenvalue equation

$$\Delta u_n = \lambda_n^2 u_n, \tag{A.1b}$$

$$\int d^2\sigma\sqrt{\gamma}u_n u_m = \delta_{nm}, \tag{A.1c}$$

$$\sum_n u_n(\sigma)u_n(\sigma') = \frac{\delta_2(\sigma-\sigma')}{\sqrt{\gamma}}, \tag{A.1d}$$

where $\delta_2(\sigma-\sigma')$ is a 2-dimensional scalar density.

The heat kernel is defined by

$$K(\sigma,\sigma';\varepsilon) = \sum_n u_n(\sigma)u_n(\sigma')e^{-\varepsilon\lambda_n^2}. \tag{A.2a}$$

We shall refer to ε as the "proper time". It follows from Eq. (A.1) that

$$\Delta K(\sigma,\sigma';\varepsilon) = -\frac{\partial}{\partial\varepsilon}K(\sigma,\sigma';\varepsilon) \tag{A.2b}$$

for $\sigma\to\sigma'$ and small ε it is convenient to expand K as

$$K(\sigma,\sigma';\varepsilon)=\frac{1}{4\pi\varepsilon}e^{-d(\sigma,\sigma')/4\varepsilon}\sum_{l=0}\varepsilon^l H_l(\sigma,\sigma') \tag{A.2c}$$

where $d(\sigma,\sigma')$ denotes the geodesic distance between σ and σ'. We assume σ is close to σ' so that there is a unique geodesic connecting them. The functions $H_n(\sigma,\sigma')$ are determined easily in the coincidence limit $\sigma=\sigma'$. If we substitute (A.2c) in (A.2b) and equate the coefficients of the various powers of ε we obtain

$$H_0(\sigma,\sigma)=1 \tag{A.3a}$$

$$H_1(\sigma,\sigma)=\tfrac{1}{6}R \tag{A.3b}$$

$$\vdots$$

It is worthwhile noting that the power of $\frac{1}{4\pi\varepsilon}$ on the right-hand side of (A.2c) is dimension dependent. It is determined from the normalization condition that in the flat n-space

$$\frac{1}{(4\pi\varepsilon)^{n/2}}e^{-|\sigma-\sigma'|^2/4\varepsilon}\xrightarrow[\varepsilon\to 0]{}\delta_n(\sigma-\sigma').$$

A.2. ***The proper time cut off***

This is a regularization scheme in a curved 2-space and has no infrared ambiguities. Following Polyakov we define the regularized $G(\sigma,\sigma';\varepsilon)$ by

$$G(\sigma,\sigma';\varepsilon)=\sum_{\lambda_n\neq 0}\frac{u_n(\sigma)u_n(\sigma')}{\lambda_n^2}e^{-\varepsilon\lambda_n^2} \tag{A.4a}$$

and immediately obtain

$$\nabla_\alpha G(\sigma,\sigma';\varepsilon)|_{\sigma=\sigma'}=\tfrac{1}{2}\nabla_\alpha G(\sigma,\sigma;\varepsilon) \tag{A.5a}$$

$$\gamma^{\alpha\beta}\nabla_\alpha\nabla'_\beta G(\sigma,\sigma';\varepsilon)|_{\sigma=\sigma'}=\gamma^{\alpha\beta}\sum_{\lambda_n\neq 0}\frac{\nabla_\alpha u_n(\sigma)\nabla_\beta u_n(\sigma)}{\lambda_n^2}e^{-\varepsilon\lambda_n^2}$$

$$=\tfrac{1}{2}\nabla^2 G(\sigma,\sigma;\varepsilon)+K(\sigma,\sigma;\varepsilon)-u_0^2. \tag{A.5b}$$

Therefore the knowledge of $G(\sigma,\sigma;\varepsilon)$ and $K(\sigma,\sigma;\varepsilon)$ are the only ingredients to determine the coincidence limits of the derivatives of G.

To evaluate $G(\sigma,\sigma;\varepsilon)$ we observe from (A.4a) that

$$\int d^2\sigma\sqrt{\gamma}\,\delta G(\sigma,\sigma;\varepsilon) = \sum_{\lambda_n\neq 0}\frac{2}{\lambda_n^2}\int d^2\sigma\sqrt{\gamma}\,\delta u_n(\sigma)u_n(\sigma)e^{-\varepsilon\lambda_n^2}$$

$$-\sum_{\lambda_n\neq 0}\frac{\delta\lambda_n^2}{\lambda_n^4}e^{-\varepsilon\lambda_n^2}$$

$$-\varepsilon\sum_{\lambda_n\neq 0}\frac{\delta\lambda_n^2}{\lambda_n^2}e^{-\varepsilon\lambda_n^2}. \tag{A.6a}$$

If we take the variation to be induced by $\delta\gamma_{\alpha\beta} = 2\delta\rho\gamma_{\alpha\beta}$ then from (A.1a)–(A.1c) it follows that

$$\delta\lambda_n^2 = -2\lambda_n^2\int d^2\sigma\sqrt{\gamma}\,\delta\rho(\sigma)u_n(\sigma)u_n(\sigma) \tag{A.6b}$$

$$\int d^2\sigma\sqrt{\gamma}\,\delta u_n(\sigma)u_n(\sigma) = -\int d^2\sigma\sqrt{\gamma}\,\delta\rho(\sigma)u_n(\sigma)u_n(\sigma). \tag{A.6c}$$

Hence (A.6a) reduces to

$$\int d^2\sigma\sqrt{\gamma}\,\delta G(\sigma,\sigma;\varepsilon) = 2\varepsilon\int d^2\sigma\sqrt{\gamma}\,\delta\rho(\sigma)(K(\sigma,\sigma;\varepsilon)-u_0^2)\xrightarrow[\varepsilon\to 0]{}\frac{1}{2\pi}\int d^2\sigma\sqrt{\gamma}\,\delta\rho(\sigma).$$

It follows that

$$G(\sigma,\sigma;\varepsilon) = \frac{1}{2\pi}\rho(\sigma) + \text{“const”}. \tag{A.7a}$$

The “constant” in Eq. (A.2a) can contain a fixed function ρ_0, an ε-dependent term and a τ-dependent part. The fixed function ρ_0 can be taken to be the constant curvature metric on the surface such that for the metrics which admit isometries—such as the rotational invariant metric on S^2, $G(\sigma,\sigma;\varepsilon)$ becomes a σ-independent constant. Hence we may write

$$G(\sigma,\sigma;\varepsilon) = \frac{1}{2\pi}(\rho(\sigma)-\rho_0(\sigma)) + g(\varepsilon,\tau). \tag{A.7b}$$

To determine the ε-dependence of $G(\sigma,\sigma;\varepsilon)$ we can substitute the short distance expansion (A.2c) into

$$G(\sigma,\sigma;\varepsilon) = \int_\varepsilon^\infty dt(K(\sigma,\sigma;t)-u_0^2),$$

write $\int_\varepsilon^\infty = \int_\varepsilon^1 + \int_1^\infty$ and ignore $\int_1^\infty$ to obtain

$$G(\sigma, \sigma; \varepsilon) = -\frac{1}{4\pi} \ln \varepsilon + \cdots .$$

Hence (A.7b) may be re-written as

$$G(\sigma, \sigma; \varepsilon) = \frac{1}{2\pi}(\rho(\sigma) - \rho_0(\sigma)) - \frac{1}{4\pi} \ln \varepsilon + g(\tau), \tag{A.7c}$$

where $g(\tau)$ may contain any other constants which may be relevant.

It can be shown that ρ_0 decouples from the gravity-dilation field equations. For this reason we have set it to zero in the text. If we do so we obtain

$$G(\sigma, \sigma; \varepsilon) = -\frac{1}{4\pi} \ln \varepsilon + \frac{1}{2\pi} \rho(\sigma) + g(\tau), \tag{A.8a}$$

$$\partial_\alpha G(\sigma, \sigma'; \varepsilon)|_{\sigma=\sigma'} = \frac{1}{4\pi} \partial_\alpha \rho(\sigma), \tag{A.8b}$$

$$\partial_\alpha \partial'_\alpha G(\sigma, \sigma'; \varepsilon)|_{\sigma=\sigma'} = \frac{1}{6\pi} \partial^2 \rho(\sigma) + \frac{e^{2\rho(\sigma)}}{4\pi\varepsilon} - e^{2\rho(\sigma)} u_0^2 . \tag{A.8c}$$

The last of these expressions seems to be regularization dependent. We shall show below that although the Pauli-Villars scheme yields the same coefficient $\frac{1}{6\pi}$ of $\partial^2 \rho$ in (A.8c) the dimensional regularization produces a value of $\frac{1}{4\pi}$. In Appendix B we shall give an exact expression for the regularized $G(\sigma, \sigma'; \varepsilon)$ on the symmetric S^2. There the regularization will be different from all the above mentioned three schemes and the value of the numerical coefficient in front of $\partial^2 \rho$ will be zero.

A.3. ***Pauli-Villars***

In this scheme one starts from the following definition of the regularized G

$$G_{\text{reg}}(\sigma, \sigma') = G(\sigma, \sigma') - \sum_r c_r G_r(\sigma, \sigma' | M_r^2) \tag{A.9a}$$

where $G(\sigma, \sigma' | M_r^2)$ satisfy

$$(\Delta + M_r^2) G(\sigma, \sigma' | M_r^2) = \frac{\delta_2(\sigma - \sigma')}{\sqrt{\gamma}} . \tag{A.9b}$$

The constants c_r and M_r^2 in (A.9a) are constrained by the requirement that if $\gamma_{\alpha\beta} = \delta_{\alpha\beta}$

then $G_{reg}(\sigma,\sigma)$, $\partial_\sigma G_{reg}(\sigma,\sigma')|_{\sigma=\sigma'}$ and $\partial_\sigma\partial'_\sigma G_{reg}(\sigma,\sigma')|_{\sigma=\sigma'}$ are regular functions of σ. To ensure this we must demand

$$\sum_r c_r = 1, \tag{A.10a}$$

$$\sum_r c_r M_r^2 = 0. \tag{A.10b}$$

Throughout this calculation we shall work in the conformal coordinates. Then Eq. (A.9b) becomes

$$(\partial^2 - M_r^2 e^{2\rho})G(\sigma,\sigma'|M_r^2) = -\delta_2(\sigma-\sigma').$$

$G(\sigma,\sigma')$ in (A.9a) satisfies a similar equation with M_r^2 replaced by zero. From (A.9a)

$$\frac{\delta G_{reg}(\sigma,\sigma')}{\delta\rho(\sigma'')} = -\sum_r c_r \frac{\delta G(\sigma,\sigma'|M_r^2)}{\delta\rho(\sigma'')}. \tag{A.11a}$$

From Eq. (A.10) it follows that

$$\frac{\delta G(\sigma,\sigma'|M_r^2)}{\delta\rho(\sigma'')} = -G(\sigma,\sigma''|M_r^2)2M_r^2 e^{2\rho(\sigma'')}G(\sigma'',\sigma'|M_r^2). \tag{A.11b}$$

Hence

$$\partial_\sigma \frac{\delta G_{reg}(\sigma,\sigma'|M_r^2)}{\delta\rho(\sigma'')} = \sum_r c_r G(\sigma,\sigma''|M_r^2)2M_r^2 e^{2\rho(\sigma'')}G(\sigma'',\sigma'|M_r^2), \tag{A.11c}$$

$$\partial_{\sigma'}\partial_\sigma \frac{\delta G_{reg}(\sigma,\sigma'|M_r^2)}{\delta\rho(\sigma'')} = \sum_r c_r \partial_\sigma G(\sigma,\sigma''|M_r^2)2M_r^2 e^{2\rho(\sigma'')}\partial_{\sigma'}G(\sigma'',\sigma'|M_r^2). \tag{A.11d}$$

Now set ρ = const. and use the Fourier transforms

$$G(\sigma,\sigma'|M_r^2) = \int \frac{d^2p}{(2\pi)^2}\frac{e^{ip\cdot(\sigma-\sigma')}}{p^2+M_r^2e^{2\rho}} \tag{A.12a}$$

$$\frac{\delta}{\delta\rho(\sigma'')}G_{reg}(\sigma,\sigma'|M_r^2)|_{\sigma=\sigma'} = \int \frac{d^2p}{(2\pi)^2}F^0(p)e^{ip\cdot(\sigma-\sigma'')}, \tag{A.12b}$$

$$\partial_\alpha\frac{\delta}{\delta\rho(\sigma'')}G_{reg}(\sigma,\sigma'|M_r^2)|_{\sigma=\sigma'} = \int \frac{d^2p}{(2\pi)^2}F_\alpha^1(p)e^{ip\cdot(\sigma-\sigma'')}, \tag{A.12c}$$

$$\partial'_\alpha\partial_\alpha\frac{\delta}{\delta\rho(\sigma'')}G_{reg}(\sigma,\sigma'|M_r^2)|_{\sigma=\sigma'} = \int \frac{d^2p}{(2\pi)^2}F^2(p)e^{2p\cdot(\sigma-\sigma'')}. \tag{A.12d}$$

The functions F^0, F^1 and F^2 can be determined by comparing Eq. (A.12) with Eq. (A.11). The results are

$$F^0(p) = \sum_r c_r 2M_r^2 e^{2\rho} \int \frac{d^2q}{(2\pi)^2} \frac{1}{(q^2 + M_r^2 e^{2\rho})[(q-p)^2 + M_r^2 e^{2\rho}]}, \tag{A.13a}$$

$$F_\alpha^1(p) = \sum_r c_r 2M_r^2 e^{2\rho} \int \frac{d^2q}{(2\pi)^2} \frac{iq_\alpha}{(q^2 + M_r^2 e^{2\rho})[(q-p)^2 + M_r^2 e^{2\rho}]}, \tag{A.13b}$$

$$F^2(p) = \sum_r c_r 2M_r^2 e^{2\rho} \int \frac{d^2q}{(2\pi)^2} \frac{q\cdot(q-p)}{(q^2 + M_r^2 e^{2\rho})[(q-p)^2 + M_r^2 e^{2\rho}]}. \tag{A.13c}$$

We shall eventually let $M_r^2 \to \infty$. The only surviving contributions will be the first few terms in a power series expansion of the functions F^0, F^1 and F^2. The results are

$$F^0(p) \xrightarrow[M_n^2\to\infty]{} \frac{1}{2\pi}, \tag{A.14a}$$

$$F_\alpha^1(p) \xrightarrow[M_n^2\to\infty]{} \frac{ip^\alpha}{4\pi}, \tag{A.14b}$$

$$F^2(p) \xrightarrow[M_n^2\to\infty]{} -\frac{1}{6\pi}p^2 + \frac{1}{2\pi}\sum_r c_r M_r^2 e^{2\rho} \ln\left(1 + \frac{\Lambda^2}{M_r^2 e^{2\rho}}\right). \tag{A.14c}$$

If we substitute these expressions in Eq. (A.11) and integrate with respect to ρ we recover the same ρ-dependencies as in (A.8).

A.4. *Dimensional regularization*

Like Pauli-Villars the dimensional regularization is an infinite flat space scheme. We start from Eq. (3.6) written in an n-dimensional space with $u_0 = 0$. Then we evaluate the response of $G(\sigma, \sigma')$ to a variation $\delta\gamma_{\alpha\beta} = 2\gamma_{\alpha\beta}\delta\rho$ under which $\delta(\sqrt{\gamma}\gamma^{\alpha\beta}) = (n-2)\sqrt{\gamma}\gamma^{\alpha\beta}\delta\rho$. We thus obtain

$$(n-2)\partial_\alpha(\delta\rho\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta)G(\sigma,\sigma') + \partial_\alpha\sqrt{\gamma}\gamma^{\alpha\beta}\partial_\beta\delta G(\sigma,\sigma') = 0.$$

Set $\gamma_{\alpha\beta} = \delta_{\alpha\beta}$ and solve for $\dfrac{\delta G}{\delta\rho}$.

$$\frac{\delta G(\sigma'',\sigma')}{\delta\rho(\sigma)} = (2-n)\partial_\alpha\Delta(\sigma''-\sigma)\partial_\alpha\Delta(\sigma-\sigma'), \tag{A.15a}$$

where Δ is defined by

$$\partial^2\Delta(\sigma-\sigma') = -\delta_{(n)}(\sigma-\sigma'). \tag{A.15b}$$

From (A.15a) it follows that

$$\partial_{\sigma''}\frac{\delta G(\sigma'',\sigma')}{\delta\rho(\sigma)}\bigg|_{\sigma''=\sigma'} = (2-n)\big(\partial'\partial_\alpha\Delta(\sigma'-\sigma)\big)\partial_\alpha\Delta(\sigma-\sigma'), \tag{A.15c}$$

$$\partial_{\sigma''}\partial_{\sigma'}\frac{\delta G(\sigma'',\sigma')}{\delta\rho(\sigma)}\bigg|_{\sigma''=\sigma'} = (2-n)\big(\partial'\partial_\alpha\Delta(\sigma'-\sigma)\big)\partial'\partial_\alpha\Delta(\sigma-\sigma'). \tag{A.15d}$$

These equations are similar to Eq. (A.11). Hence we again use the notation of Eq. (A.12). and obtain the following expressions for the functions F^0, F^1, F^2.

$$F^0(p) = (2-n)\int\frac{d^nq}{(2\pi)^n}\frac{q\cdot(q-p)}{(q^2+m^2)[(q-p)^2+m^2]}, \tag{A.16a}$$

$$F^1_\alpha(p) = (2-n)\int\frac{d^nq}{(2\pi)^n}\frac{q\cdot(q-p)i(q-p)_\alpha}{(q^2+m^2)[(q-p)^2+m^2]}, \tag{A.16b}$$

$$F^2(p) = (2-n)\int\frac{d^nq}{(2\pi)^n}\frac{[q\cdot(q-p)]^2}{(q^2+m^2)[(q-p)^2+m^2]}, \tag{A.16c}$$

where m^2 is an infrared regulator. It is straightforward to calculate the pole parts of the integrals as $(2-n)\to 0$. The result is

$$F^0(p) \xrightarrow[n\to 2]{} \frac{1}{2\pi}, \tag{A.17a}$$

$$F^1_\alpha(p) \xrightarrow[n\to 2]{} -\frac{1}{4\pi}ip_\alpha, \tag{A.17b}$$

$$F^2(p) \xrightarrow[n\to 2]{} -\frac{1}{4\pi}p_\alpha^2. \tag{A.17c}$$

These expressions reproduce the same ρ-dependencies for $G(\sigma,\sigma)$ and $\partial G(\sigma,\sigma')|_{\sigma=\sigma'}$, as in (A.8a) and (A.8b). However the coefficient of $\partial^2\rho$ in (A.8c) becomes $\frac{1}{4\pi}$ instead of $\frac{1}{6\pi}$.

Appendix B Exact Expression for Regularized $G(\sigma,\sigma')$ on S^2

On a S^2 of radius b and with a symmetric metric the eigenfunctions of Δ are given by the spherical harmonics $Y_{l,m}(\theta,\varphi)$

$$\Delta Y_{lm}(\theta,\varphi) = \frac{l(l+1)}{a^2} Y_{lm}(\theta,\varphi). \tag{B.1}$$

Hence

$$G_\alpha(\theta,\varphi;\theta',\varphi') = \frac{1}{\alpha} + \sum_{l\neq 0} \sum_{m=-l}^{l} \frac{Y^*_{lm}(\theta,\varphi) Y_{lm}(\theta',\varphi')}{l(l+1)} \tag{B.2}$$

where α is a constant "gauge parameter". Using the well-known formulae

$$\sum_{m=-l}^{l} Y^*_{lm}(\theta,\varphi) Y_{lm}(\theta',\varphi') = \frac{2l+1}{4\pi} P_l(\cos\omega), \tag{B.3a}$$

$$\cos\omega = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\varphi-\varphi'), \tag{B.3b}$$

we get

$$G_\alpha(\theta,\varphi;\theta',\varphi') = \frac{1}{\alpha} + \frac{1}{4\pi} \sum_{l\neq 0} \frac{2l+1}{l(l+1)} P_l(\cos\omega). \tag{B.4}$$

Now if we write

$$\frac{2l+1}{l(l+1)} = \int_0^1 dt(t^{l-1} + t^l) \quad l \geq 1 \tag{B.5a}$$

and use the following generating function of P_l

$$\frac{1}{\sqrt{1-2t\cos\omega + t^2}} = \sum_{l=0}^{\infty} t^l P_l(\cos\omega) \tag{B.5b}$$

then (B.4) becomes

$$G_\alpha(\theta,\varphi;\theta',\varphi') = \frac{1}{\alpha} + \frac{1}{4\pi} \int_0^1 dt \left(\frac{1}{t}+1\right)\left(\frac{1}{\sqrt{1-2t\cos\omega+t^2}} - 1\right). \tag{B.6}$$

It is clear from (B.6) that in the coincidence limit $\cos\omega = 1$ the integral diverges at $t = 1$ as in $(1-t)$.

Regularization

We define the regularized G_α by

$$G_\alpha(\sigma,\sigma';\varepsilon) = \frac{1}{\alpha} + \frac{1}{4\pi} \int_0^{1-\varepsilon} dt \left(\frac{1}{t}+1\right)\left(\frac{1}{\sqrt{1-2t\cos\omega+t^2}} - 1\right) \tag{B.7}$$

where $\varepsilon \to 0$ is an ultraviolet regulator (*not* to be confused with the heat kernel ε). It is not too hard to perform the t-integration. The result is

$$G_\alpha(\sigma, \sigma'; \varepsilon) = \text{const.} - \frac{1}{2\pi} \ln[\varepsilon + \sqrt{\varepsilon^2 + 2(1-\varepsilon)(1-\cos\omega)}] + 0(\varepsilon). \tag{B.8}$$

Now we can study two limits

(i) In the coincidence limit $\sigma = \sigma'$ and $\cos\omega = 1$ hence

$$G_\alpha(\sigma, \sigma, \varepsilon) = -\frac{1}{2\pi} \ln \varepsilon + \text{const.} \tag{B.9}$$

(ii) If $\sigma \neq \sigma'$ we can set $\varepsilon = 0$ to obtain

$$G_\alpha(\sigma, \sigma', \varepsilon = 0) = -\frac{1}{4\pi} \ln\left(\frac{1-\cos\omega}{2}\right) + \text{const.} \tag{B.10}$$

It is easy to show that

$$\frac{1-\cos\omega}{2} = \frac{|Z - Z'|^2/2b^2}{\left(1 + \frac{|Z|^2}{2b^2}\right)\left(1 + \frac{|Z'|^2}{2b^2}\right)} \tag{B.11}$$

where $Z = \sqrt{2}b \cot\theta/2 e^{i\phi}$ defines the stereographic projection. Substitution of (B.11) into (B.10) reproduces Eq. (3.9) of the text.

To conclude we examine the double derivatives of (B.10) in the coincidence limit

$$\partial_Z \partial'_{\bar{Z}} G_\alpha(\sigma, \sigma'; \varepsilon) = \frac{1-\varepsilon}{4\pi\varepsilon^2} \frac{1}{b^2} e^{2\rho} \tag{B.12}$$

where

$$e^{2\rho} = \frac{1}{\left(1 + \frac{|Z|^2}{2b^2}\right)^2}.$$

The right-hand side of (B.12) is different from the prediction of all three regularization schemes we discussed in Appendix A. This indicates that there is no regularization independent significance in the double derivatives of $G(\sigma, \sigma')$ in the coincidence limit.

References

1. For reviews, see: J. Schwarz, "Superstrings: The first fifteen years" (World Scientific, Singapore, 1985); For the heterotic superstrings, see: D. Gross, J. Havey, E. Martinec, and R. Rohm, *Nucl. Phys.* **B256** (1985) 253 and *Nucl. Phys.* **B267** (1986) 75.
2. P. Candelas, G. Horowitz, A. Strominger, and E. Witten, *Nucl. Phys.* **B258** (1985) 47.
3. J. Scherk, and J. H. Schwarz, *Nucl. Phys.* **B81** (1974) 118; *Phys. Lett.* **52B** (1974) 347; T. Yoneya, *Nuovo Cimento Lett.* **8** (1973) 951; *Prog. Theor. Phys.* **51** (1974) 1907; R. I. Neopomechie, *Phys. Rev.* **32D** (1985) 3201; J. Gross, and E. Witten, *Nucl. Phys.* **B277** (1986) 1.
4. C. Lovelace, *Phys. Lett.* **135B** (1984) 75; C. G. Callen, D. Friedan, E. J. Martinec, and M. J. Perry, *Nucl. Phys.* **B262** (1985) 593; A. Sen, *Phys. Rev. Lett.* **55** (1985) 1846; S. P. de Alwis, *Phys. Rev.* **D34** (1986) 3760.
5. E. S. Fradkin, and A. A. Tseytlin, *Phys. Lett.* **158B** (1985) 316; *Nucl. Phys.* **B61** (1985) 1.
6. The number of references in this connection is quite large. See in particular: M. T. Grisaru, and D. Zanon, "Sygma model corrections to the Einstein-Hilbert action", preprint 86; A. sen, *Phys. Rev.* **D32** (1985) 2102; *Phys. Rev. Lett.* **55** (1985) 1846; C. Hull, and P. K. Townsend, *Nucl. Phys.* **B274** (1986) 349; G. Shore, "A local renormalization group equation, diffeomorphisms, and conformal invariance in sigma models", preprint BUTP-86/16.
7. D. Friedan, *Phys. Lett.* **45** (1980) 1057; B. Z. Freedman, and P. K. Townsend, *Nucl. Phys.* **B177** (1981) 443.
8. L. Alvarez-Gaumé, D. Z. Freedman, and S. Mukhi, *Ann. Phys. (N.Y.)* **134** (1981) 85.
9. M. T. Grisaru, A. E. M. Van de Van, and D. Zanon, "Four loop-function for the $N = 1$ and $N = 2$ supersymmetric nonlinear sigma models in two dimensions", preprint 1986; *Nucl. Phys.* **B277** (1986) 409.
10. B. S. De Witt, *Dynamical Theory of Groups and Fields* (Blackie, London, 1965).
11. C. G. Callan, I. R. Kebanov, and M. J. Perry, *Nucl. Phys.* **B272** (1986) 111.
12. D. Mumford, *Tata Lectures on Theta I* (Birkhäuser, Boston 1983) p. 36.
13. S. P. de Alwis, *Phys. Lett.* **168B** (1986) 59.
14. H. Arfai, *Nucl. Phys.* **B112** (1976) 256; J. Plchinski, *Commun. Math. Phys.* **104** (1986) 37.
15. A. H. Polyakov, *Phys. Lett.* **103B** (1981) 207.
16. See C. G. Callan *et al.* in Ref. 4 and Ref. 11.
17. A. M. Namazie, K. S. Narain, and H. Sarmadi, *Phys. Lett.* **177B** (1986) 329; S. Hamidi, C. Vafa, "Interactions on orbifolds", *Nucl. Phys.* **B279** (1987) 465, see the appendix.
18. K. Fujikawa, *Phys. Rev. Lett.* **44** (1980) 1733; P. van Nieuwenhuizen, "Critical dimensions and supersymmetric measures", preprint ITP-5B-86-84.
19. For an interesting new suggestion see the recent papers by C. G. Callan, C. Lovelace, C. R. Nappi, and S. A. Yost, "String loop corrections to beta functions", preprint PUPT-1027; A. Abouelsaud, C. G. Callan, C. R. Nappi and S. A. Yost, Princeton preprint, September 1986.

    The approach of these authors is based on an earlier proposal by W. Fischler and L. Susskind, *Phys. Lett.* **173B** (1986) 493.

Volume 135B, number 5,6 PHYSICS LETTERS 16 February 1984

ON KALUZA–KLEIN COSMOLOGY

S. RANDJBAR-DAEMI [1]
International Centre for Theoretical Physics, Trieste, Italy
and Institute for Theoretical Physics, University of Bern, Bern, Switzerland

Abdus SALAM
International Centre for Theoretical Physics, Trieste, Italy
and Imperial College, London, England

and

J. STRATHDEE
International Centre for Theoretical Physics, Trieste, Italy

Received 16 November 1983

A generalization of Friedmann–Robertson–Walker cosmology to $4+K$ dimensions is considered. The space–time manifold $R^1 \times S^3 \times S^K$ is characterized by two time-dependent scales, $R(t)$ and $a(t)$. The equations of motion for R and a are derived from the $4+K$-dimensional Einstein action supplemented by a one-loop thermal term, corresponding to a gas of non-interacting scalar particles. It is shown that in the approximation when $T < 1/a$ the equations of motion admit a solution in which the internal space has a constant radius a while the external $R(t)$ evolves in the usual manner.

1. Einstein equations. The idea of an internal space whose symmetries correspond to the observed internal symmetries of low energy physics has received much attention, recently. It is assumed in this picture that space–time has more than four dimensions and that all but four of them are highly curved and compactified so as to be unresolvable with the energies available at present. The relevant distance (measured in units of Planck length) would be inversely correlated with the gauge coupling constants in the theory.

As with other questions involving inaccessibly small distance scales it seems appropriate to look for cosmological implications of the Kaluza–Klein hypothesis. Were the compact dimensions always compact and always so much more curved than the others? If there is a temporal evolution of the distance scales, is it compatible with the observed constancy of gauge coupling parameters like the fine structure constant? We shall demonstrate that it is indeed possible to maintain a mixed internal scale while the external radius evolves in the conventional manner.

To study the cosmological aspects [1] of the Kaluza–Klein hypothesis we shall begin by setting up a straightforward generalization of the homogeneous Friedmann, Roberston, Walker cosmologies for space–times with extra dimensions. To be specific we shall assume the geometry $R^1 \times S^3 \times S^K$ where R^1 corresponds to the time dimension, S^3 to the "large" space dimensions and S^K to the "small" or internal space dimensions. Instead of the sphere S^3 one could choose an open space of constant curvature and, instead of the sphere S^K one could choose a compact quotient space G/H of K dimensions. In any case the metric tensor could be presented in the form

$$g_{MN} = \begin{pmatrix} -1 & & \\ & R(t)^2 \tilde{g}_{mn}(x) & \\ & & a(t)^2 \tilde{g}_{\mu\nu}(y) \end{pmatrix}, \quad (1.1)$$

[1] Supported by Schweizerischer National Funds.

Reprinted with permission from *Phys. Lett.* 135B (1984) 388–392
© 1984 Elsevier Science Publishers B.V.

where $\tilde{g}_{mn}(x)$ and $g_{\mu\nu}(y)$ represent metrics on the three- and K-dimensional spaces, respectively. The sizes of these two spaces are characterized by time-dependent scale factors, $R(t)$ and $a(t)$.

The metric (1.1) admits a number of isometries. For the case $R^1 \times S^3 \times S^K$, the symmetry group of the group of the spatial section is $O(4) \times O(K+1)$. The stress tensor, whose form would be dictated by Einstein's equations, must have the same invariances. In the coordinate system where the metric is given by (1.1), the stress tensor is

$$T_{MN} = \begin{pmatrix} \rho & & \\ & p g_{mn} & \\ & & p' g_{\mu\nu} \end{pmatrix}, \tag{1.2}$$

where the energy density ρ and pressures p, p' may depend on time, but not on the space coordinates x^m, y^μ.

The Einstein equations for the scale factors R and a in terms of ρ, p and p' can be derived from the action functional,

$$\Gamma = \int d^{4+K}x(-g)^{1/2}[-G^{-1}(\mathbf{R} + \bar{\Lambda})] + \Gamma_1(g), \tag{1.3}$$

where $\mathbf{R}$ denotes the $4+K$-dimensional Ricci scalar, $\bar{G}$ is the coupling parameter and $\bar{\Lambda}$ is a cosmological constant. The source term, Γ_1, represents the contributions of matter fields, vacuum fluctuations, etc. Its variation defines the stress tensor,

$$\delta\Gamma_1 = \int d^{4+K}x(-g)^{1/2}\tfrac{1}{2}\delta g^{MN} T_{MN}. \tag{1.4}$$

By substituting the $O(4) \times O(K+1)$ invariant ansatz (1.1) into (1.3), one can reduce this action to the form

$$\Gamma = \int dt\, \mathcal{L}(R, a),$$

where the lagrangian, $\mathcal{L}$, is given by

$$\mathcal{L} = -(\Omega_3\Omega_K/\bar{G}) \times [6\dot{R}^2/R^2 + 6K(\dot{R}/R)\dot{a}/a + K(K-1)\dot{a}^2/a^2] - U. \tag{1.5}$$

Here Ω_3 and Ω_K denote the spatial volumes,

$$\Omega_3 = 2\pi^2 R^3, \quad \Omega_K = (2\pi)^{(K+1)/2} a^K/\Gamma((K+1)/2) \tag{1.6}$$

and U is an effective "potential". The construction of this potential will be discussed more fully in the following section. It incorporates a classical gravity term arising from $\mathbf{R} + \Lambda$ [and displayed in the first bracket of (1.7)] plus terms in the second bracket which arise from Γ_1. The latter has a (one-loop) quantum part plus a thermal part depending on entropy, S

$$U = (\Omega_3\Omega_K/\bar{G})[-6/R^2 - K(K-1)/a^2 + \bar{\Lambda}] + (\sigma_K\Omega_3/a^4 + \tau S^{4/3}R + \zeta m). \tag{1.7}$$

Here σ_K, τ and ζ are numerical parameters which can be computed. The approximations which go into the form of the entropy (S)-dependent term in (1.7) will be discussed in the next section, but we believe that the given expression should be valid in the regime

$$R \gg 1/T \gg a, \tag{1.8}$$

where the temperature T is given by

$$T = \partial U/\partial S \sim S^{1/3}/R. \tag{1.9}$$

The time evolution of the Kaluza–Klein universe is governed by the Euler–Lagrange equations derived from (1.5). We now show that for an appropriate choice of the parameter $\bar{\Lambda}$ in U, it is possible to obtain a solution where the internal radius is constant ($\dot{a} = 0$) while the large radius evolves as in standard model cosmology.

The two equations of motion simplify, on setting $\dot{a} = 0$, to the following:

$$\partial_t(R\dot{R}) - \tfrac{1}{2}\dot{R}^2 = \tfrac{1}{12}\bar{G}(R^2/\Omega_3\Omega_K)R\,\partial U/\partial R, \tag{1.10a}$$

$$R^{-1}\partial_t(R^2\dot{R}) - \dot{R}^2 = (\bar{G}/6K)(R^2/\Omega_3\Omega_K)a\partial U/\partial a. \tag{1.10b}$$

The "energy" integral corresponding to these equations reduces, when $\dot{a} = 0$, to

$$E = -(\Omega_3\Omega_K/\bar{G})(6\dot{R}^2/R^2) + U. \tag{1.11}$$

The compatibility of (1.10a), (1.10b) and (1.11) requires the algebraic relation

$$R\,\partial U/\partial R - (2/K)a\partial U/\partial a = E - U. \tag{1.12}$$

On taking $E = \zeta m$, this condition reduces to

$$f(a) = df/da = 0, \tag{1.13a}$$

where [‡1]

[‡1] Eqs. (13a) are the same as those of Candelas and Weinberg [2] for the static case considered by them ($R = \infty$).

Volume 135B, number 5,6 PHYSICS LETTERS 16 February 1984

$$f(a) = (\Omega_K/\bar{G})[-K(K-1)/a^2 + \bar{\Lambda}] + \sigma/a^4 . \quad (1.13b)$$

Determining $\bar{\Lambda}$ and a in terms of $\bar{G}$ from (1.13a) and substituting the results into (1.10b), we obtain,

$$\partial_t(R\dot{R}) = -1 , \quad (1.14)$$

which is solved by

$$R^2 = t(t_0 - t) , \quad (1.15)$$

whre t_0, the time of collapse, is given by (1.11), i.e. $t_0^2 = (\tau/3\pi^2)(\bar{G}/\Omega_K)S^{4/3}$.

The expression (1.15) for $R^2(t)$ gives the conventional Friedmann–Robertson–Walker cosmology for the case when the standard three-curvature parameter k equals unity. In the general case, the $6/R^2$ term in the first bracket in (1.7) reads $6k/R^2$ and the right-hand side of (1.14) equals $(-k)$, with appropriate changes in (1.15).

To summarize, if $R \gg 1/T > a$, and if $\bar{\Lambda}$ and a are determined in terms of $\bar{G}$ from (1.13a), we find $\dot{a}(t) = 0$, while $R(t)$ evolves in the conventional manner.

2. The free energy. In order to have a meaningful expression for the components of the stress tensor, it is necessary to derive them from a free energy function, $F(R, a, T)$. From F one obtains the energy as a function of the entropy, $S = -\partial F/\partial T$, by the usual Legendre transform

$$U = F + TS = \Omega_3\Omega_K\rho(R, a, S) . \quad (2.1)$$

If the pressures p and p' of eq. (1.2) are defined by the thermodynamic identity,

$$d(\Omega_3\Omega_K\rho) = T dS - \Omega_K p\, d\Omega_3 - \Omega_3 p'\, d\Omega_K , \quad (2.2)$$

then as is well-known the conservation of energy, $\nabla_M T^M{}_N = 0$, implies conservation of entropy. It follows that the Einstein equations deriving from (1.5) require that S be independent of time. This was tacitly assumed in section 1.

Since we have no a priori idea of what equations of state would be appropriate for Kaluza–Klein cosmology, we shall proceed on the basis of a simple model. We shall suppose that the density and pressures are due to a gas of non-interacting, spinless particles in thermal equilibrium. The free energy of such a system can be expressed by a one-loop integral,

$$\beta F = \tfrac{1}{2} \ln \mathrm{De}\, t(-\Box + \mu^2) , \quad (2.3)$$

where μ is a mass parameter and $\Box$ denotes the laplacian on the compact manifold $S^1 \times S^3 \times S^K$. This manifold is characterized by three radii, $\beta/2\pi = 1/2\pi T$, R and a. The expression (2.3) is given formally by a triple sum over the eigenvalues of $-\Box + \mu^2$,

$$\beta F = \frac{1}{2} \sum_{r=-\infty}^{\infty} \sum_{m,n=0}^{\infty} D_{mn} \times \ln[(2\pi r/\beta)^2 + m(m+2)/R^2 + n(n+K-1)/a^2 + \mu^2] , \quad (2.4)$$

where D_{mn} is a multiplicity factor. It is equal to the dimension of the $O(4) \times O(K+1)$ representation $(m, 0) \times (n, 0, ..., 0)$ in Gel'fand Zetlin notation,

$$D_{mn} = (m+1)^2 \times (2n+K-1)(n+K-2)!/(K-1)!n! . \quad (2.5)$$

The sum (2.4) is of course divergent. But it is easy to regularize [3]. For this purpose write

$$\ln X = \frac{d}{ds} X^s|_{s=0} = \frac{d}{ds}\left(\frac{1}{\Gamma(-s)} \int_0^\infty dt\, t^{s-1} \exp(-tX)\right)_{s=0}$$

for each term in the sum. After some rearrangements one obtains an expression for the finite part

$$\beta F = \frac{d}{ds}\left(\frac{1}{2\Gamma(-s)} \int_0^\infty dt\, t^{-s-1} \exp(-t\mu^2) \times \sigma_1(4\pi^2 t/\beta^2)\sigma_3(t/R^2)\sigma_K(t/a^2)\right)_{s=0} , \quad (2.6)$$

where the functions σ_1, σ_3 and σ_K are defined, respectively, by

$$\sigma_1(4\pi^2 t/\beta^2) = \sum_{-\infty}^{\infty} \exp[-t(2\pi r/\beta)^2] ,$$

$$\sigma_3(t/R^2) = \sum_0^\infty (m+1)^2 \exp[-m(m+2)t/R^2] ,$$

$$\sigma_K(t/a^2) = \sum_0^\infty [(2n+K-1)(n+K-2)!/(K-1)!n!] \times \exp[-n(n+K-1)t/a^2] . \quad (2.7)$$

These sums converge and define analytic functions on the half-plane Re $t > 0$. At $t = 0$ they are singular,

$$\sigma_1 \sim t^{-1/2}, \quad \sigma_3 \sim t^{-3/2}, \quad \sigma_K \sim t^{-K/2}.$$

To evaluate the integral (2.6) it is necessary to make approximations to the σ_i. The most important observation in this connection [4] is that these functions are closely related to the Jacobi Theta functions [5]. In particular,

$$\theta_3(0|iu/\pi) = \sum_{-\infty}^{\infty} \exp(-r^2 u)$$

$$= \sqrt{\pi/u} \sum_{-\infty}^{\infty} \exp[-(\pi r)^2 u^{-1}] . \qquad (2.8)$$

The two series on the right-hand side provide highly convergent expansions for the σ_i around $u = \infty$ and around $u = 0$ [‡2], respectively. In particular, the zero-temperature limit corresponds to the term

$$\sigma_1 \simeq (4\pi t/\beta^2)^{-1/2} , \qquad (2.9)$$

and gives

$$F(R, a, T = 0) = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{1}{4\sqrt{\pi}\Gamma(-s)}\right.$$

$$\left. \times \int_0^{\infty} \mathrm{d}t\, t^{-s-3/2} \exp(-t\mu^2)\sigma_3(t/R^2)\sigma_K(t/a^2)\right).$$

In a similar fashion, the flat space limit, $R \to \infty$, of the energy density, $F/2\pi^2 R$, corresponds to

$$\sigma_3 \simeq \tfrac{1}{4}\sqrt{\pi} R^3 t^{-3/2} . \qquad (2.10)$$

The effective potential of section 1 was obtained by assuming $R \gg \beta > a$, while the mass parameter μ for the scalar particles is small (more precisely $\mu < T$). To derive it, use (2.9) and (2.11) to obtain,

$$F \simeq \frac{R^3}{16} \frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{1}{\Gamma(-s)} \int_0^{\infty} \mathrm{d}t\, t^{-s/3}\right.$$

$$\left. \times \left(1 + 2\sum_1^{\infty} \exp(-r^2\beta^2/4t)\right)\sigma_K(t/a^2)\right]_{s=0}$$

$$= \frac{R^3}{16}\left[\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{a^{-4-2s}}{\Gamma(-s)} \int_0^{\infty} \mathrm{d}u\, u^{-s-3}\sigma_K(u)\right)_{s=0}\right.$$

$$\left. - 2\int_0^{\infty} \mathrm{d}t\, t^{-3} \sum_1^{\infty} \exp(-r^2\beta^2/4t)\sigma_K(t/a^2)\right] . \qquad (2.11)$$

The "s-regularization" can be safely removed in the temperature dependent terms, which converge at $t = 0$. It can also be removed in the zero-temperature term if K is odd, since the integral involving $\sigma_K(u)$ is analytic at $s = 0$ in that case. (For even K one obtains $\sim a^{-4}$ ln a in the limit $s = 0$.) Thus

$$F \simeq \Omega_3(\sigma_K a^{-4} - \tfrac{1}{90}\pi^2\beta^{-4} + \ldots) , \qquad (2.12)$$

where

$$\sigma_K = -\frac{1}{32\pi^2}\int_0^{\infty} \mathrm{d}u\, u^{-3}\sigma_K(u) . \qquad (2.13)$$

In obtaining the second term of (2.12) we have taken $\sigma_K = 1$, in effect discarding the exponentially small terms $\sim \exp(-\beta/a)$. (Otherwise stated, only the internal space zero modes are retained in the sum over states.) Details of these various approximations will be discussed elsewhere.

The one-loop contribution to the energy, used in section 1, is then obtained by taking the Legendre transform of (2.12)

$$U \simeq \sigma_K \Omega_3/a^4 + \tau S^{4/3}/R , \qquad (2.14)$$

where

$$\tau = \tfrac{3}{4}(45/4\pi^4)^{1/3} . \qquad (2.15)$$

[If there are other species of particles, in addition to scalars which circulate around the loop (for example gravitons, gauge particles, or spinors) the numerical values of the parameters σ_K and τ will change.]

3. Conclusions. It is interesting to confirm that the

[‡2] This seems to be true for the odd-dimensional spheres. For the even-dimensional spheres one probably has to employ asymptotic series $\sigma_K(u - K/2(a + \beta u + \gamma u 2 + \ldots)$, near $u = 0$.

Kaluza–Klein cosmology does admit of a time-independent internal radius a (consistent with lack of variability of gauge couplings with time), in an approximation to the free energy where $R \gg 1/T > a$, while $\bar{\Lambda}$ and a are suitably correlated by (1.13a). For the regime of temperatures where our *precise* approximation for the thermal contribution to the free energy, $(S^{4/3}/R)$, may be expected to break down – for example for planckian temperatures, $T \gtrsim 1/a$ – phase transitions [2] in the viour of a (and R) may occur. These will considered elsewhere.

The authors would like to thank Professor L Van Hove for a discussion.

References

[1] P.G.O. Freund, Nucl. Phys. B209 (1982) 146;
T. Appelquist and A. Chodos, Phys. Rev. Lett. 50 (1983) 141;
E. Alvarez and M. Belen Gavela, Phys. Rev. Lett. 51 (1983) 931;
M. Rubin and B.D. Roth, Phys. lett. 127B (1983) 55;
Nucl. Phys. B226 (1983) 44;
T. Appelquist, A. Chodos and E. Meyers, Yale preprint YPT 83-04 (1983);
Q. Shafi and C. Wetterich, Phys. Lett. 129B (1983) 387.
[2] P. Candelas and S. Weinberg, Texas preprint UTTG-8-83 (1983).
[3] A. Salam and J.A. Strathdee, Nucl. Phys. B90 (1975) 203.
[4] C.M. Mukku, J.A. Magpantay and W.A. Sayed, Ann. Phys. (NY) 145 (1983) 27.
[5] A. Erdelyi et al., eds., Higher transcendental functions (McGraw-Hill, New York, 1953) Vol. II.

5. Condensed Matter and Biology

The papers collected in this final section represent something of a departure from Salam's earlier interests. Almost all the previous papers (save only perhaps the very first) deal with the properties of elementary particles. In recent years, however, his interests have ranged more widely, encompassing in particular fundamental problems in condensed matter physics and in biology.

In the field of superconductivity, the standard model that has served the test of time is the theory of Bardeen, Cooper and Schrieffer (BCS), which provides an excellent description of the metallic superconductors. Like the unified electroweak theory, the BCS model is a spontaneously broken gauge theory. The symmetry in question is ordinary electromagnetic gauge invariance. The underlying theory is of course fully gauge-invariant, but the ground state is not. The phonon-mediated interaction between electrons leads to a condensation of 'Cooper pairs', pairs of electrons of equal and opposite momentum and spin. Thus the photon acquires an effective mass and magnetic fields are expelled from the superconductor — the Meissner effect.

In the mid-80s the world of superconductivity was revolutionized by the discovery by Georg Bednorz and Alex Müller of a quite different class of non-metallic compounds exhibiting superconductivity up to remarkably high temperatures, well over 100 K. The underlying mechanism is not yet well understood; the BCS model provides no explanation for their behaviour.

One of the important characteristics of the materials in question is their anisotropic structure, characterized by two-dimensional copper-oxide layers. The electron motion is thus largely confined to two dimensions. It is therefore not surprising that theorists have sought an explanation for the curious properties of high-temperature superconductors in two-dimensional models.

There is a remarkable class of models in (2+1)-dimensional space-time, in which *unbroken* gauge symmetry is associated with massive quanta and short-range forces. These are the Chern-Simons theories, of which the paradigm is described by the Lagrangian density

$$\mathcal{L} = -\frac{1}{4e^2} F_{\lambda\mu} F^{\lambda\mu} - \frac{\nu}{2\pi} \varepsilon^{\lambda\mu\rho} A_\lambda \partial_\mu A_\rho + j^\mu A_\mu \,, \tag{38}$$

with $F_{\lambda\mu} = \partial_\lambda A_\mu - \partial_\mu A_\lambda$. Provided the current j^μ is conserved, this Lagrangian is gauge-invariant up to a total divergence. In two spatial dimensions, the gauge coupling e has the dimensions of mass, while the Chern-Simons coupling ν is dimensionless. If space is compactified, ν is restricted to being integral by the requirement of gauge-invariance of the exponential of the action, $\exp\left(i \int \mathcal{L}\, d^3x\right)$.

Seifallah Randjbar-Daemi, Abdus Salam and John Strathdee studied some of the implications of a theory in which the electrons are coupled to an effective Chern-Simons gauge field, in particular computing the finite temperature corrections and showing that the Meissner effect present at zero temperature persists at finite temperature, though as yet it is unclear how realistic the model can be.

The same authors have also studied generalized spin systems. Systems of coupled spins, such as the Heisenberg model of a ferromagnet, have played an important role in many-body theory. In such models, each point n of a lattice is associated with a spin operator, $\mathbf{S}^{(n)}$, whose components obey the SU(2) commutation relations. Randjbar-Daemi, Salam and Strathdee considered extended models in which at each lattice site there is an operator belonging to a representation of a larger symmetry group. They derived many interesting features of the models, showing in particular that their continuum limits are σ-models.

The other subject represented in this final section is fundamental theoretical biology, in particular the question of why it is that most biological molecules appear in nature in only one of two mirror-image forms. How did this chirality originate? Can it be related to the lack of mirror symmetry in the weak interactions?

The weak neutral force mediated by the Z particle does give rise to a small energy difference between the L and D isomeric forms of molecules such as amino acids. A number of authors have calculated this energy difference and found that the L form generally has lower energy by an amount of order 10^{-19} eV. This is a very tiny advantage, and it is unclear that it would have been enough to ensure the overwhelming dominance of the L forms in biological materials — though of course one must bear in mind the enormously long time, of the order of billions of years, over which such an effect might have operated.

Salam proposed a radical solution to this problem, namely that a phase transition is involved, with a Bose condensation below some critical temperature, as in the BCS theory of superconductivity. It seems unlikely, however, that the critical temperature could be high enough to allow this condensation to occur on the primitive Earth, so he suggested that the original selection may have occurred elsewhere. This is certainly an intriguing theory, with the merit of being fairly readily testable.

It is fitting that this selection should end, as it began, with papers demonstrating Salam's penetrating insight, originality and wide interests.

Nuclear Physics B340 (1990) 403–447
North-Holland

CHERN–SIMONS SUPERCONDUCTIVITY AT FINITE TEMPERATURE

S. RANDJBAR-DAEMI, A. SALAM and J. STRATHDEE

International Centre for Theoretical Physics, Trieste, Italy

Received 2 October 1989
(Revised 12 February 1990)

A simple gauge theory discussed recently in the literature as a model of high-temperature superconductors is examined. The model contains a Maxwell field and a Chern–Simons field coupled to fermions in (2 + 1)-dimensional spacetime. This model has been shown to exhibit a kind of Meissner effect at zero temperature, which originates in the 1-loop mixing between the two gauge fields. We use a euclidean effective action formulation to show that the effect persists at finite temperatures. Although a long-range magnetic type interaction arises at non-zero temperatures, in competition with the finite range forces which dominate the zero temperature interaction, the effect varies smoothly with temperature.

1. Introduction

Gauge symmetries in (3 + 1)-dimensional spacetime are normally associated with long-range forces such as electromagnetism. In the appropriate circumstances, however, the gauge symmetry can be spontaneously broken, in which cases the associated force has a finite range. This mechanism is invoked in the BCS theory of superconductivity where a weakly attractive force between electrons gives rise to a condensate of Cooper pairs. The BCS ground state is not invariant with respect to the electromagnetic gauge transformations and the photon, in effect, acquires a mass. Magnetic fields are consequently expelled from the superconductor. This mechanism is non-perturbative in the sense that the formation of quasi-bound states, the Cooper pairs, is an essential feature. On the other hand, it has been known for some years that gauge symmetries in (2 + 1)-dimensional spacetime are not necessarily associated with long range forces [1]. Here it is consistent with unbroken gauge symmetry to have massive gauge quanta. In the lagrangian formulation the mass term is represented by a Chern–Simons density,

$$\mathscr{L} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} - \frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}A_\lambda\partial_\mu A_\rho + j^\mu A_\mu . \tag{1.1}$$

This density is invariant, up to a total derivative with respect to the U(1) gauge

© Elsevier Science Publishers B.V.

transformations, $A_\mu \to A_\mu + \partial_\mu \Lambda$, provided the current is conserved, $\partial_\mu j^\mu = 0$. The coupling parameter, e^2, has the dimensions of mass (in natural units) while the second coupling parameter, ν, is dimensionless. It is an elementary exercise to show that the effective current–current interaction, due to the exchange of a gauge quantum, is given by

$$\frac{1}{2}\left(\frac{1}{k^2+\left((\nu/\pi)e^2\right)^2}\right)\left[e^2 j_\mu(-k)j^\mu(k) - \frac{i\pi}{\nu}\varepsilon^{\mu\lambda\rho}j_\mu(-k)k_\lambda j_\rho(k)\right]$$

$$+\frac{1}{2}\frac{1}{k^2}\frac{i\pi}{\nu}\varepsilon^{\mu\lambda\rho}j_\mu(-k)k_\lambda j_\rho(k)\,. \tag{1.2}$$

This indicates that the force has both finite and long-range components. Both of these forces have a parity violating piece with strength π/ν. (The long-range component in fact describes an instantaneous interaction of the currents – see eq. (2.24).)

The parameter ν is arbitrary if the 2-dimensional space has euclidean topology. However, if the space is compactified then it becomes possible to consider topologically non-trivial gauge transformations under which the Chern–Simons term in eq. (1.1) is not invariant. Then, in order to maintain invariance of the phase, $\exp(i\int\mathscr{L})$, it becomes necessary to restrict ν to be an integer [2].

The relevance of (2 + 1)-dimensional gauge theory to the description of high-temperature superconductors [3] is suggested by the possibility that the layered structure of copper oxide crystals causes the electron motion to be effectively 2-dimensional. One could then consider (1.1) as an effective lagrangian, in which the coupling parameter, e^2, is obtained by averaging over the thickness of the layers

$$e^2 \sim \alpha/\delta\,,$$

where $\alpha \sim 1/137$ is the usual fine structure constant and $\delta \sim 10^{-7}$ cm is a measure of thickness. The parameter, ν, is not so easy to interpret – it would have to be generated by some parity violating effect in the underlying crystal structure. In this paper we shall examine a simple model which has been considered recently by a number of authors [4, 8]. It contains a fundamental Chern–Simons field, a_μ, which couples to the "electrons" exactly like the electromagnetic A_μ. This Chern–Simons field should be viewed as a kind of vector excitation of the underlying microscopic dynamics. It can be shown that the introduction of a fundamental Chern–Simons gauge field coupled to 2-dimensional fermions is mathematically equivalent to modifying the statistics of these fermions [5] – changing them into what are called "anyons" [6]. It is known that such particles, although otherwise free, can exhibit an effective interaction which is caused by their fractional statistics [7]. This is

demonstrated by eliminating the Chern–Simons field as an independent dynamical variable [4]. Here we shall adopt the point of view that it is more practical to keep the Chern–Simons field among the independent variables and treat it as a gauge field on the same footing as the electromagnetic field. The electrons will therefore obey normal Fermi statistics. It is then possible to carry out perturbative calculations by expanding in powers of π/ν, which is to be regarded as a small parameter. Using this approach, the authors of refs. [4, 8] were able to show that, at the level of 1-loop quantum corrections, an effective Chern–Simons type of term involving the electromagnetic potential, A_μ, is induced along with the usual ground-state polarization effects. In other words, the photon develops a mass in the lowest order of perturbation theory. Specifically, the effective lagrangian includes the term

$$\frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}(A_\lambda + a_\lambda)\partial_\mu(A_\rho + a_\rho), \tag{1.3}$$

along with other bilinear terms involving the electric and magnetic field strengths. This phenomenon can be interpreted as a kind of 2-dimensional superconductivity with a Meissner effect [4, 8] (finite penetration depth) and a supercurrent which is proportional to the Chern–Simons electric-type field strength.

In sect. 2 we review the model and describe the 1-loop computation in terms of an effective action formulation. Although this simple model may be somewhat artificial, it is at least arguable that it may be typical of a class of microscopic theories which incorporate P and T violation and which lead to an induced electromagnetic Chern–Simons term. We have studied the effective interaction and shown that there are two independent short-range forces, corresponding to the exchange of massive quanta. In the static limit one of these ranges is identical to the penetration depth found in refs. [4, 8]. The other is generally shorter.

To test this theory further, we have made a finite temperature computation which is described in sect. 3 and appendix C. The results are used in sect. 4 to obtain the effective current–current interaction at finite temperature. We find that, generally, there are three independent components to the magnetic interaction, two of finite range, corresponding to the exchange of massive quanta and, in addition a long-range component which vanishes in the zero-temperature limit. The relative strengths of these interactions vary smoothly with temperature. In the low-temperature regime we obtain a formula for the temperature dependence of the parameter, λ, identified as "penetration depth" in refs. [4, 8]

$$\frac{1}{\lambda^2} = \frac{e^2 n_e}{m}(1 - 4y\,e^{-y}) + \frac{4\pi n_e}{\nu^2} y\,e^{-y}$$

where m and n_e denote the electron mass and number density, respectively, and

$y = \pi n_e/2\nu mT \sim 10^4$ K$/\nu T \gg 1$. Notice that $1/\lambda$ grows with T. Further, we find that the long-range component dominates at high temperatures. The two finite penetration depths appear to vary smoothly with temperature and, at least in the low and high regimes studied in sect. 4, cannot diverge to infinity. There is no indication of critical behaviour. However, the shorter of these goes to zero at a finite temperature which suggests an instability. (See note added in proof.)

The failure of the present model in perturbation theory to provide a "realistic" description of high-temperature superconductivity, notwithstanding, we believe that it may still be possible to develop a more successful version based on the Chern–Simons idea*. We have therefore set out our calculations in some detail in the appendices.

2. Effective action

The model to be considered in 2 + 1 dimensions involves a gas of non-relativistic electrons coupled to two independent gauge fields. The latter are represented by the electromagnetic potential, A_μ, and the Chern–Simons potential, a_μ. The lagrangian for this system is

$$\mathscr{L} = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} - \frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}a_\lambda\partial_\mu a_\rho + \psi^+ i\nabla_0\psi - \frac{1}{2m}\nabla_k\psi^+\nabla_k\psi\,, \qquad (2.1)$$

where the electron field, ψ, is a two-component spinor whose covariant derivatives are given by

$$\nabla_\mu\psi = (\partial_\mu - iA_\mu - ia_\mu)\psi\,,$$

$$\nabla_\mu\psi^+ = (\partial_\mu + iA_\mu + ia_\mu)\psi^+.$$

The lagrangian of Hosotani and Chakravarty [8] includes a magnetic moment term $\sim F_{12}\psi^+\sigma_3\psi$, but we shall ignore this for the present. The system described by eq. (2.1) can be quantized in the standard way and Feynman rules given for the perturbative development in powers of the coupling parameters e^2 and π/ν. In principle, one could construct an effective action functional to any given order by evaluating irreducible graphs. However, since our main interest lies in finding effective field equations for the two gauge fields which are accurate to 1-loop order, we shall simply integrate out the electron field to obtain

$$\Gamma(A,a) = -\frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} - \frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}a_\lambda\partial_\mu a_\rho + \frac{1}{i}\ln\det G(A+a)\,, \qquad (2.2)$$

* For example, it is widely believed that the finite temperature behaviour of anyon superconductors will be of the Kosterlitz–Thouless type. If this turns out to be the case, then the perturbative description given here will not be adequate.

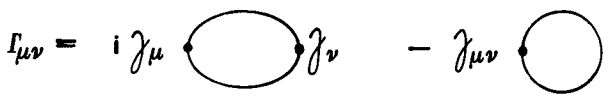


Fig. 1. One-loop ground-state polarization tensor.

where $G(A+a)$ denotes the electron Green function in a specified background. The functional determinant in eq. (2.2) gives the complete 1-loop contribution to Γ. Higher loop contributions would require the evaluation of graphs involving internal A- and a-lines and so would be proportional to positive powers of the small parameters e^2 and π/ν. We shall therefore ignore them. Because of the assumed minimal couplings, the 1-loop term is a functional of the sum, $A_\mu + a_\mu$. The introduction of non-minimal terms in eq. (2.1) would destroy this simplicity*.

The next simplification that we shall make is to assume that the fields A_μ and $a'_\mu = a_\mu - a^0_\mu$ are small and we shall evaluate Γ up to second order in these small quantities. Here, the "background" a^0_μ represents the ground-state expectation value and corresponds to a uniform Chern–Simons magnetic field. We shall take

$$a^0_0 = a^0_2 = 0, \qquad a^0_1 = -\frac{x_2}{l^2}, \tag{2.3}$$

where l^2 is a constant to be determined. This background enters both the electron propagator and the vertices. It causes the electron energy levels to be quantized in units of $1/ml^2$ (Landau levels) and the resulting electron propagator $G(a^0)$ is neither translation nor rotation invariant. However, it should be recognized that the breaking of these continuous symmetries is only a gauge artifact since the background field strength, $f^0_{\mu\nu} = \partial_\mu a^0_\nu - \partial_\nu a^0_\mu$, is invariant. Covariance with respect to combined rotations, translations and suitable background gauge transformations will be maintained (see appendix A). Gauge-invariant quantities such as the ground-state polarization tensors will turn out to have the correct spacetime covariance. On the other hand, with respect to discrete symmetries, such as the reflection $x_1 \to -x_1$, $x_2 \to x_2$, the background magnetic field is not invariant and these symmetries are broken.

The leading terms in the development of the 1-loop effective action can be expressed in terms of ground-state expectation values of the time-ordered products

* At the end of this section we include the contribution due to an electron magnetic moment, see eq. (2.21).

of electron current operators

$$\Gamma_{(1)} = \int \mathrm{d}^3x\, \Gamma^{\mu}(x)\big(A_{\mu}(x) + a'_{\mu}(x)\big)$$

$$+ \tfrac{1}{2}\int \mathrm{d}^3x\, \mathrm{d}^3x'\, \Gamma^{\mu\nu}(x,x')\big(A_{\mu}(x) + a'_{\mu}(x)\big)\big(A_{\nu}(x') + a'_{\nu}(x')\big) + \ldots, \quad (2.4)$$

where (see fig. 1)

$$\Gamma^{\mu}(x) = -\langle j^{\mu}(x)\rangle,$$

$$\Gamma^{\mu\nu}(x,x') = i\langle Tj^{\mu}(x)j^{\nu}(x')\rangle - \langle j^{\mu\nu}(x,x')\rangle, \quad (2.5)$$

with the current operators given by

$$j_0(x) = \psi^{+}\psi,$$

$$j_k(x) = \frac{i}{2m}\big(\psi^{+}\nabla_k\psi - \nabla_k\psi^{+}\psi\big),$$

$$j_{0\mu}(x,x') = 0,$$

$$j_{kl}(x,x') = \frac{1}{m}\delta_{kl}\delta_3(x-x')\psi^{+}\psi, \quad (2.6)$$

in which ∇_k indicates the covariant derivative with the background connection a^0_{μ}. (We are using the minkowskian metric for which $j_0 = -j^0$, $j_k = j^k$).

The electron Green function,

$$G_{\sigma\sigma'}(x,x') = i\langle T\psi_{\sigma}(x)\psi^{+}_{\sigma'}(x')\rangle \quad (2.7)$$

is obtained by solving the equation

$$\left(i\partial_0 + \frac{1}{2m}\nabla_k^2\right)G(x,x') = -\delta_3(x-x'), \quad (2.8)$$

subject to appropriate boundary conditions. It can be expressed by the integral

$$G(x,x') = -\int_{C_N}\frac{\mathrm{d}\omega}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k_1}{2\pi}\mathrm{e}^{ik_1(x_1-x'_1)-i\omega(t-t')}$$

$$\times\frac{1}{l}\sum_n\frac{v_n[(x_2/l)+k_1l]\,v_n^{*}[(x'_2/l)+k_1l]}{\omega-\varepsilon_n}, \quad (2.9)$$

where the functions v_n, $n = 0, 1, \ldots$, are orthonormalized harmonic oscillator wave functions. They satisfy the eigenvalue equations

$$\frac{1}{2m}\left[\left(k_1 + \frac{x_2}{l^2}\right)^2 - \partial_2^2\right] v_n = \varepsilon_n v_n ,$$

$$\varepsilon_n = \left(n + \tfrac{1}{2}\right)\frac{1}{ml^2} . \tag{2.10}$$

The choice of contour C_N for the integration over frequencies is dictated by the boundary conditions which define the ground state, viz.

$$G(x, x') \sim \begin{cases} \sum\limits_{n<N} e^{-i\varepsilon_n(t-t')}, & t < t', \\ \sum\limits_{n\geqslant N} e^{-i\varepsilon_n(t-t')}, & t > t', \end{cases}$$

where N is an integer which counts the number of filled Landau levels. Hence the contour C_N must pass below the poles at $\omega = \varepsilon_n$ for $n = 0, 1, \ldots, N-1$ and above all the rest, $n \geqslant N$. Using this propagator it is straightforward to compute the quantities (2.5)

$$\Gamma^0(x) = i \operatorname{Tr} G(x, x')\big|_{x'=x,\, t'=t+0},$$

$$\Gamma^k(x) = \frac{1}{2m}(\nabla_k - \nabla_k')\operatorname{Tr} G(x, x')\Big|_{x'=x,\, t'=t+0}, \tag{2.11}$$

$$\Gamma^{00} = i \operatorname{Tr} G(x, x')G(x', x),$$

$$\Gamma^{k0}(x, x') = \frac{1}{2m}\operatorname{Tr}(\nabla_k G(x, x')G(x', x) - G(x, x')\nabla_k G(x', x)),$$

$$\Gamma^{kl}(x, x') = -\frac{i}{4m^2}\operatorname{Tr}(\nabla_k G(x, x')\nabla_l' G(x', x) - G(x, x')\nabla_k \nabla_l' G(x', x))$$

$$+ \nabla_l' G(x, x')\nabla_k G(x', x) - \nabla_k \nabla_l' G(x, x')G(x', x))$$

$$-\frac{i}{m}\delta_{kl}\delta_3(x - x')\operatorname{Tr} G(x, x')\Big|_{x'=x,\, t'=t+0}. \tag{2.12}$$

The linear components (2.11) are given by

$$\Gamma^0(x) = \frac{N}{\pi l^2}, \qquad \Gamma^k(x) = 0, \tag{2.13}$$

The polarization components (2.12), being gauge independent, must have translation and rotation symmetry. Further, because of current conservation their Fourier components can be expressed in terms of three independent invariants*,

$$\begin{aligned}
\tilde{\Gamma}_{00} &= \boldsymbol{k}^2 \Pi_0 , \\
\tilde{\Gamma}_{j0} &= -\omega k_j \Pi_0 - i\varepsilon_{jl} k_l \Pi_1 , \\
\tilde{\Gamma}_{jl} &= \omega^2 \delta_{jl} \Pi_0 + i\varepsilon_{jl}\omega \Pi_1 + \left(k^2 \delta_{jl} - k_j k_l\right)\Pi_2 ,
\end{aligned} \tag{2.14}$$

where $\omega = k^0 = -k_0$. The general structure of the invariants Π_0, Π_1 and Π_2 is considered in appendix B. They are functions of ω^2 and k^2 but here we need only their threshold values

$$\Pi_0 = \frac{N}{\pi} m l^2, \qquad \Pi_1 = \frac{N}{\pi}, \qquad \Pi_2 = -\frac{N^2}{\pi m} . \tag{2.15}$$

These are zeroth-order terms in expansions in powers of $(kl)^2$ and $(\omega m l^2)^2$.

The results (2.13) and (2.15) define the behaviour of the 1-loop effective action for small fluctuations of the gauge potentials around the ground state values $A_\mu = 0$, $a_\mu = a_\mu^0$. For a consistent determination of the ground state it is necessary to satisfy the effective equations of motion

$$\frac{\delta \Gamma}{\delta A_\mu} = n_e \delta_0^\mu, \qquad \frac{\delta \Gamma}{\delta a_\mu} = 0, \tag{2.16}$$

where n_e denotes an external uniform charge density representing an average density of compensating positive charge. (It is assumed that only the electromagnetic potential is coupled to this charge.) The functional derivatives in eq. (2.16) are to be evaluated at $A = 0$, $a = a^0$ where the non-vanishing components are

$$\frac{N}{\pi l^2} = n_e, \qquad -\frac{\nu}{\pi}\frac{1}{l^2} + \frac{N}{\pi l^2} = 0, \tag{2.17}$$

respectively. The first term in the latter equation represents the classical contribution $\delta S/\delta a_0$ evaluated at the point (2.3). It follows that the ground state parame-

* Where $\varepsilon_{12} = -\varepsilon_{21} = 1$. The three-dimensional permutation symbol is defined such that $\varepsilon^{012} = 1 = -\varepsilon_{012}$, etc. Current conservation and rotation invariance alone would allow four independent invariants but the lagrangian (2.1) is also PT invariant.

ters, N and l^2 are fixed by the consistency requirement

$$N = \nu \quad \text{and} \quad l^2 = \frac{\nu}{\pi n_e}. \tag{2.18}$$

This means, in particular, that the coupling parameter, ν, which was introduced in the classical lagrangian (2.1), must be an integer, positive or negative. (In the above treatment we have assumed that ν is positive. Consistency for the case of negative ν requires merely a change of sign in eq. (2.3) and subsequent formulae.)

Having dealt with the linear term in the effective action and its role in determining the ground-state parameters, we can now combine the polarization terms (2.15) with the classical terms to define an effective lagrangian which is bilinear in the gauge potentials A_μ and a'_μ.

$$\mathscr{L}_{\text{eff}} = \frac{1}{2e^2}\left(E_j^2 - B^2\right) - \frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}a'_\lambda\partial_\mu a'_\rho + \frac{1}{2}\left(\frac{\nu}{\pi}\right)^2\frac{m}{n_e}(E_j + e_j)^2$$

$$-\frac{1}{2}\left(\frac{\nu}{\pi}\right)^2\frac{\pi}{m}(B+b)^2 + \frac{\nu}{2\pi}\varepsilon^{\lambda\mu\rho}(A_\lambda + a'_\lambda)\partial_\mu(A_\rho + a'_\rho), \tag{2.19}$$

where the electric and magnetic field strengths are defined by

$$E_j = \partial_0 A_j - \partial_j A_0, \qquad B = \partial_1 A_2 - \partial_2 A_1,$$

$$e_j = \partial_0 a'_j - \partial_j a'_0, \qquad b = \partial_1 a'_2 - \partial_2 a'_1.$$

In obtaining eq. (2.19) we have discarded higher derivative terms. Hence this lagrangian should be used for the description of fields which are slowly varying in the sense

$$l\partial_j E \ll E, \qquad ml^2\partial_0 E \ll E, \tag{2.20}$$

etc. At this point it may be remarked that the inclusion of an electron magnetic moment term $-\mu_e B\psi^+\sigma_3\psi$ in the original lagrangian would give rise to a higher derivative term in eq. (2.19)

$$\frac{1}{2}\left(\frac{\nu}{\pi}\right)^2\frac{m}{n_e}\mu_e^2(\partial_j B)^2. \tag{2.21}$$

With $\mu_e = 1/2m$ this term would become significant at longer wavelengths than those which are excluded by eq. (2.20) but we shall nevertheless discard it along with all the other higher derivative terms.

From the effective lagrangian (2.19), we can obtain expressions for the gauge field Green functions. The procedure for doing this is straightforward in principle

but the computations are somewhat lengthy because of the mixing and they are not very illuminating. Some of the details are given in appendix D. The main result is that the energy spectrum has two branches given by the zeros of the function

$$\Delta(\omega^2,k^2)=\left(\frac{\pi}{\nu e^2}\right)^2\left[\left\{(\omega^2-k^2)\Pi_0-\frac{\nu}{\pi}e^2\Pi_1\right\}^2-(\omega^2-k^2)\left(\frac{\nu}{\pi}\right)^2\left(1-\frac{\pi}{\nu}\Pi_1+e^2\Pi_0\right)^2\right.$$

$$\left.+k^2(\Pi_0+\Pi_2)\left\{(\omega^2-k^2)\Pi_0-\left(\frac{\nu}{\pi}\right)^2e^2(1+e^2\Pi_0)\right\}\right], \qquad (2.22)$$

which is quadratic in ω^2. The roots are real and positive if $e^2\Pi_0>0$. In the limit $k\to 0$ they define the rest energies

$$\mu_\pm^2=\left(\frac{\nu}{2\pi}\right)^2\frac{1}{\Pi_0^2}\left[\left(1-\frac{\pi}{\nu}\Pi_1+e^2\Pi_0\right)\pm\sqrt{\left(1-\frac{\pi}{\nu}\Pi_1-e^2\Pi_0\right)^2+4e^2\Pi_0}\right]^2. \qquad (2.23)$$

(It may be remarked that in the special situation, $\Pi_2=-\Pi_0$, the lagrangian (2.19) becomes Lorentz invariant as does the function (2.22). In this case the rest energies (2.23) can be interpreted as rest masses. In practice, as we shall see, $\Pi_2\ll\Pi_0$, and the system is far from relativistic.)

The static interaction, mediated by the exchange of zero frequency gauge quanta, is governed by the zeros of the function $\Delta(0,k^2)$ which is quadratic in k^2. There are two roots, $k_\pm^2$, which are both negative, corresponding to forces of finite range. In the relativistic case these ranges are given by the Compton wavelengths $\mu_\pm^{-1}$, but in general they are given by quite complicated expressions. The finite ranges of the static interaction are a manifestation of the Meissner effect in this system. They define two independent "penetration depths". Since our main concern in this paper is to study the temperature dependence of this effect, we postpone further discussion to sect. 4, where the finite temperature version of the function $\Delta(0,k^2)$ is obtained.

3. Thermodynamic potential

To obtain a description of the system at finite temperature we shall apply the methods of euclidean field theory and the grand canonical ensemble [9]. That is, we shall make the substitution $t\to -i\tau$ and regard the real parameter τ as a coordinate on a circle of circumference $\beta=1/T$. Bosonic fields are required to be single-valued (periodic) on this circle while fermions are double-valued (antiperiodic). The time components of vectors such as A_μ and J_μ are likewise redefined

$$A_0\to iA_\tau, \qquad J_0\to iJ_\tau$$

and the conjugate electron field ψ^+ is replaced by $\bar{\psi}$. We obtain the euclidean action

$$S^{\mathrm{E}} = \int_0^\beta \mathrm{d}\tau \int \mathrm{d}^2x \left[\frac{1}{4e^2} F_{\mu\nu}F_{\mu\nu} + \frac{i\nu}{2\pi} \varepsilon_{\lambda\mu\rho} a_\lambda \partial_\mu a_\rho + \bar{\psi} \nabla_\tau \psi + \frac{1}{2m} \nabla_k \bar{\psi} \nabla_k \psi - \mu \bar{\psi} \psi \right], \tag{3.1}$$

where $\varepsilon_{\tau 12} = 1$, and the covariant derivatives are defined in the natural way

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \qquad \mu, \nu = \tau, 1, 2,$$

$$\nabla_\mu \psi = (\partial_\mu - iA_\mu - ia_\mu)\psi,$$

$$\nabla_\mu \bar{\psi} = (\partial_\mu + iA_\mu + ia_\mu)\bar{\psi}.$$

It is important to keep in mind that the components A_τ and a_τ are imaginary and that $\bar{\psi}$ is not the hermitian conjugate of ψ. Note that the coefficient of the Chern–Simons term is now purely imaginary.

The euclidean action (3.1) contains a new parameter, μ, the chemical potential. It will be determined as a function of temperature and the other parameters in eq. (3.1) by the conditions of thermal equilibrium to be described.

A euclidean effective action can be defined in close analogy to the minkowskian case discussed in sect. 2. One proceeds by introducing external sources I_μ and J_μ for the gauge potentials A_μ and a_μ, respectively. Define the functional $W(I, J)$ by path integration over periodic boson and antiperiodic fermion fields

$$\mathrm{e}^{-W(I,J)} = \int (\mathrm{d}A\, \mathrm{d}a\, \mathrm{d}\psi\, \mathrm{d}\bar{\psi}) \exp\left[-S^{\mathrm{E}} - \int \mathrm{d}^3x (I_\mu A_\mu + J_\mu a_\mu) \right].$$

The averaged fields are defined by functional differentiation

$$A_\mu = \frac{\delta W}{\delta I_m}, \qquad a_\mu = \frac{\delta W}{\delta J_\mu}$$

and are to be treated as independent variables. The euclidean effective action, Γ^{E}, is then given by the Legendre transform of W

$$\Gamma^{\mathrm{E}}(A, a) = W(I, J) - \int \mathrm{d}^3x (I_\mu A_\mu + J_\mu a_\mu).$$

It can be expressed as a loop expansion in which the zeroth term is the classical

euclidean action

$$\Gamma^{E}(A,a) = S^{E}(A,a) + \Gamma^{E}_{(1)}(A,a) + \dots . \tag{3.2}$$

The successive terms are represented by irreducible graphs. In the following we shall obtain the leading 1-loop terms in a weak field expansion around the constant Chern–Simons background (2.3)

$$\Gamma^{E}_{(1)} = \int d^3x\, \Gamma^{E}_{\mu}(x)\big(A_{\mu}(x) + a'_{\mu}(x)\big)$$

$$+ \tfrac{1}{2}\int d^3x\, d^3x'\, \Gamma^{E}_{\mu\nu}(x,x')\big(A_{\mu}(x) + a'_{\mu}(x)\big)\big(A_{\nu}(x) + a'_{\nu}(x')\big) + \dots . \tag{3.3}$$

The coefficient functions in this case are expressed in terms of τ-ordered correlation functions

$$\Gamma^{E}_{\mu}(x) = -\big\langle j_{\mu}(x)\big\rangle ,$$

$$\Gamma^{E}_{\mu\nu}(x,x') = -\big\langle T_{\tau} j_{\mu}(x) j_{\nu}(x')\big\rangle - \big\langle j_{\mu\nu}(x,x')\big\rangle , \tag{3.4}$$

where the euclidean current components are given by

$$j_{\tau}(x) = i\bar{\psi}\psi ,$$

$$j_{k}(x) = -\frac{i}{2m}\big(\bar{\psi}\nabla_{k}\psi - \nabla_{k}\bar{\psi}\psi\big) ,$$

$$j_{\tau\mu}(x,x') = 0 ,$$

$$j_{kl}(x,x') = -\frac{1}{m}\delta_{kl}\delta_{3}(x-x')\bar{\psi}\psi . \tag{3.5}$$

To evaluate the correlators (3.4) we need the electron thermal Green function [9]

$$\mathcal{G}(x,x') = -\big\langle T_{\tau}\psi(x)\bar{\psi}(x')\big\rangle . \tag{3.6}$$

This Green function is obtained by solving the equation

$$\Big(\partial_{\tau} - \frac{1}{2m}\nabla_{k}^{2} - \mu\Big)\mathcal{G}(x,x') = -\delta_{3}(x-x') , \tag{3.7}$$

subject to the requirement of antiperiodicity under the translation $\tau \to \tau + \beta$. It

can be expanded in normal modes

$$\mathcal{G}(x,x') = \frac{1}{\beta}\sum_s \int \frac{dk_1}{2\pi} e^{ik_1(x_1-x_1')-i\zeta_s(\tau-\tau')} \frac{1}{l}\sum_n \frac{v_n[(x_2/l)+k_1 l]\, v_n^*[(x_2'/l)+k_1 l]}{i\zeta_s - \varepsilon_n + \mu}, \tag{3.8}$$

where the functions v_n are the same harmonic oscillator eigenfunctions that appeared in sect. 2. The discrete frequencies, ζ_s, are half-integer multiples of $2\pi/\beta$

$$\zeta_s = \left(s+\frac{1}{2}\right)\frac{2\pi}{\beta}, \qquad s \in \mathbb{Z}. \tag{3.9}$$

In terms of the thermal Green function (3.8), the correlators (3.4) are given by

$$\begin{aligned}
\Gamma_\tau^{\mathrm{E}}(x) &= -i \operatorname{Tr} \mathcal{G}(x,x')\big|_{x'=x,\tau'=\tau+0}, \\
\Gamma_k^{\mathrm{E}}(x) &= \frac{i}{2m}(\nabla_k - \nabla_k')\operatorname{Tr}\mathcal{G}(x,x')\Big|_{x'=x,\tau'=\tau+0}, \\
\Gamma_{\tau\tau}^{\mathrm{E}}(x,x') &= -\operatorname{Tr}(\mathcal{G}(x,x')\mathcal{G}(x',x)), \\
\Gamma_{k\tau}^{\mathrm{E}}(x,x') &= \frac{1}{2m}\operatorname{Tr}(\nabla_k \mathcal{G}(x,x')\mathcal{G}(x',x) - \mathcal{G}(x,x')\nabla_k \mathcal{G}(x',x)), \\
\Gamma_{kl}^{\mathrm{E}}(x,x') &= -\frac{1}{4m^2}\operatorname{Tr}(\nabla_k \mathcal{G}(x,x')\nabla_l' \mathcal{G}(x',x) - \nabla_k \nabla_l' \mathcal{G}(x,x')\mathcal{G}(x',x) \\
&\qquad + \nabla_l' \mathcal{G}(x,x')\nabla_k \mathcal{G}(x',x) - \mathcal{G}(x,x')\nabla_k \nabla_l' \mathcal{G}(x',x)) \\
&\quad + \frac{1}{m}\delta_{kl}\delta_3(x-x')\operatorname{Tr}\mathcal{G}(x,x')\Big|_{x'=x,\tau'=\tau+0},
\end{aligned} \tag{3.10, 3.11}$$

where all covariant derivatives refer to the background Chern–Simons field (2.3).

Some details of the computation are discussed in appendix C. The result for the components (3.10) is given by

$$\begin{aligned}
\Gamma_\tau^{\mathrm{E}}(x) &= \frac{-i}{\pi l^2}\sum_n (e^{\beta(\varepsilon_n-\mu)}+1)^{-1}, \\
\Gamma_k^{\mathrm{E}}(x) &= 0.
\end{aligned} \tag{3.12}$$

The polarization components (3.11) can be represented by Fourier expansions

$$\Gamma^{\mathrm{E}}_{\mu\nu}(x,x') = \frac{1}{\beta}\sum_s \int \frac{\mathrm{d}^2k}{(2\pi)^2} \mathrm{e}^{ik\cdot(x-x')-i\omega_s(\tau-\tau')}\tilde{\Gamma}^{\mathrm{E}}_{\mu\nu}(\omega_s,\boldsymbol{k}), \tag{3.13}$$

where, since the currents are periodic under $\tau \to \tau + \beta$, the frequencies are integer multiples of $2\pi/\beta$

$$\omega_s = \frac{2\pi s}{\beta}, \qquad s \in \mathbb{Z}. \tag{3.14}$$

Again, because of current conservation and rotation symmetry, there are three independent invariants

$$\begin{aligned}
\tilde{\Gamma}^{\mathrm{E}}_{\tau\tau} &= \boldsymbol{k}^2 \Pi^{\mathrm{E}}_0, \\
\tilde{\Gamma}^{\mathrm{E}}_{j\tau} &= \omega_s k_j \Pi^{\mathrm{E}}_0 + \varepsilon_{jl}k_l \Pi^{\mathrm{E}}_1, \\
\tilde{\Gamma}^{\mathrm{E}}_{jl} &= \omega_s^2 \delta_{jl}\Pi^{\mathrm{E}}_0 + \varepsilon_{jl}\omega_s \Pi^{\mathrm{E}}_1 + \left(\boldsymbol{k}^2\delta_{jl} - k_j k_l\right)\Pi^{\mathrm{E}}_2.
\end{aligned} \tag{3.15}$$

The invariants Π^{E}_0, Π^{E}_1 and Π^{E}_2 are functions of ω_s^2 and $\boldsymbol{k}^2$ and we shall be interested in their behaviour near $\omega_s = k = 0$. It turns out that Π^{E}_1, Π^{E}_2 are regular there but Π^{E}_0 has a simple pole at $\boldsymbol{k}^2 = 0$ in the amplitude with $\omega_s = 0$. The results, for $\omega_s = 0$, are

$$\begin{aligned}
\Pi^{\mathrm{E}}_0 &= \frac{1}{\boldsymbol{k}^2}\frac{\beta}{4\pi l^2}\sum_n \operatorname{sech}^2 \tfrac{1}{2}\beta(\varepsilon_n - \mu) \\
&\quad + \frac{ml^2}{\pi}\sum_n \left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1} - \frac{\beta}{8\pi}\sum_n (2n+1)\operatorname{sech}^2 \tfrac{1}{2}\beta(\varepsilon_n-\mu), \\
\Pi^{\mathrm{E}}_1 &= \frac{1}{\pi}\sum_n \left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1} - \frac{1}{8\pi}\frac{\beta}{ml^2}\sum_n (2n+1)\operatorname{sech}^2 \tfrac{1}{2}\beta(\varepsilon_n-\mu), \\
\Pi^{\mathrm{E}}_2 &= \frac{1}{\pi m}\sum_n (2n+1)\left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1} - \frac{1}{16\pi}\frac{\beta}{m^2l^2}\sum_n (2n+1)^2\operatorname{sech}^2 \tfrac{1}{2}\beta(\varepsilon_n-\mu).
\end{aligned} \tag{3.16}$$

The results (3.12) and (3.16) define the behaviour of the 1-loop contribution to the euclidean effective action for weak long-wave fluctuations of the gauge potentials around the equilibrium values $A_\mu = 0$, $a_\mu = a^0_\mu$. For a consistent deter-

mination of the equilibrium configuration it is necessary to satisfy the euclidean equations of motion

$$\frac{\delta \Gamma^{\mathrm{E}}}{\delta A_\mu} = -in_{\mathrm{e}}\delta_{\mu\tau}, \qquad \frac{\delta \Gamma^{\mathrm{E}}}{\delta a_\mu} = 0, \tag{3.17}$$

where, as before, n_{e} represents an external uniform charge density. The functional derivatives are to be evaluated at $A_\mu = 0$, $a_\mu = a_\mu^0$. According to eq. (3.12) the non-vanishing components are

$$\frac{1}{\pi l^2}\sum_n \left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1} = n_{\mathrm{e}},$$

$$\frac{\nu}{\pi}\frac{1}{l^2} - \frac{1}{\pi l^2}\sum_n \left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1} = 0, \tag{3.18}$$

where the first term in the latter equation represents the 0-loop contribution. The equilibrium value of the magnetic length, l, is therefore given by

$$l^2 = \frac{\nu}{\pi n_{\mathrm{e}}}, \tag{3.19}$$

independently of the temperature. On the other hand, the chemical potential, μ, is determined by the equation

$$\nu = \sum_n \left(\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1\right)^{-1}, \tag{3.20}$$

which indicates that the Chern–Simons parameter ν is to be identified with the thermal average of the electron occupation numbers. (At zero temperature the right-hand side of eq. (3.20) reduces to the number of occupied Landau levels in the ground state.)

The coefficients (3.16) determine the bilinear part of the euclidean effective action for slowly varying fields

$$|k| \ll \frac{1}{l}, \qquad 2\pi|s| \ll \frac{\beta}{ml^2}. \tag{3.21}$$

If we restrict our considerations to fields which are independent of τ then the euclidean action can be interpreted as a thermodynamic potential

$$\Gamma^{\mathrm{E}} = \beta\Omega \tag{3.22}$$

with

$$\Omega = \int \mathrm{d}^2x \Big[\frac{1}{2e^2}\left(E_j^2 + B^2\right) + \frac{i\nu}{\pi} a_\tau b + \tfrac{1}{2}(E_j + e_j)\Pi_0^{\mathrm{E}}(E_j + e_j) + \tfrac{1}{2}\Pi_2^{\mathrm{E}}(B+b)^2 - i\Pi_1^{\mathrm{E}}(A_\tau + a_\tau)(B+b)\Big], \quad (3.23)$$

where, because of its pole at $k^2 = 0$, Π_0^{E} is to be understood as an integral operator. The functional (3.23) can be used to evaluate the thermal averages of interaction energies associated with static current distributions. For example, the thermal averaged fields A_μ and a'_μ associated with an external electromagnetic current distribution, $I_\mu(x)$, are obtained by solving the equations

$$\frac{\delta\Omega}{\delta A_\mu(x)} = -I_\mu(x), \qquad \frac{\delta\Omega}{\delta a_\mu(x)} = 0.$$

The value of the thermodynamic potential is then given by the integral

$$\Omega_I = \tfrac{1}{2}\int \mathrm{d}^2x I_\mu A_\mu, \quad (3.24)$$

which is a bilinear functional of the currents. Some aspects of this computation will be examined in sect. 4.

Next, we give the leading terms in the low-temperature expansions of the expressions (3.12) and (3.16). The characterization "low" is taken here to mean that the temperature is small in comparison with the level spacing, or

$$\beta \gg ml^2 = \frac{\nu}{\pi}\frac{m}{n_{\mathrm{e}}}. \quad (3.25)$$

The first step is to obtain an approximate expression for the chemical potential by solving eq. (3.20). To this end, we write

$$\mathrm{e}^{\beta(\epsilon_n - \mu)} = w^{(\rho - 1/2) - n}, \quad (3.26)$$

where

$$w = \mathrm{e}^{-\beta/ml^2}, \qquad \rho = \mu ml^2. \quad (3.27)$$

For small w it is possible to expand the right-hand side of eq. (3.21) in fractional powers of w

$$\sum_n \left(1 + w^{(\rho - 1/2) - n}\right)^{-1} = \left[\rho + \tfrac{1}{2}\right] + \sum_{r \geq 1} (-)^r \frac{w^{r\delta} - w^{r(1-\delta)} - w^{r(\rho + 1/2)}}{1 - w^r}, \quad (3.28)$$

where $[\rho + \frac{1}{2}]$ denotes the positive integer defined by

$$\rho + \tfrac{1}{2} = [\rho + \tfrac{1}{2}] + \delta, \qquad 0 < \delta < 1. \tag{3.29}$$

If w is sufficiently small then eq. (3.20) takes the form

$$\nu = [\rho + \tfrac{1}{2}] - w^{\delta} + w^{1-\delta} + w^{(\rho+1/2)+\delta} + \ldots$$

and is solved in leading order by

$$[\rho + \tfrac{1}{2}] = \nu, \qquad \delta = \tfrac{1}{2} + \frac{w^{\nu}}{2 \ln w}.$$

The chemical potential is therefore given by

$$\mu = \frac{\nu}{ml^2} - \frac{1}{2\beta} e^{-\beta\nu/ml^2} + \ldots. \tag{3.30}$$

On substituting this result into eq. (3.26) we obtain

$$e^{\beta(\varepsilon_n - \mu)} = \left\{ \exp\left[\frac{\beta}{ml^2}(n + \tfrac{1}{2} - \nu) \right] \right\} \left(1 + \tfrac{1}{2} e^{-\beta(\nu+1/2)/ml^2} + \ldots \right). \tag{3.31}$$

It is now a simple exercise to find the leading terms in the expressions (3.16)

$$\Pi_0^{\mathrm{E}} = \frac{1}{k^2} \frac{2\beta}{\pi l^2} e^{-\beta/2ml^2} + ml^2 \frac{\nu}{\pi} \left(1 - \frac{2\beta}{ml^2} e^{-\beta/2ml^2} \right),$$

$$\Pi_1^{\mathrm{E}} = \frac{\nu}{\pi} \left(1 - \frac{2\beta}{ml^2} e^{-\beta/2ml^2} \right),$$

$$\Pi_2^{\mathrm{E}} = \frac{\nu^2}{\pi m} \left[1 + \frac{2}{\nu^2} e^{-\beta/2ml^2} - \frac{2\beta}{ml^2} \left(1 + \frac{1}{4\nu^2} \right) e^{-\beta/2ml^2} \right]. \tag{3.32}$$

These results will be applied in sect. 4 to find the temperature dependence of the penetration depth.

For completeness we now give the high-temperature approximations for these amplitudes. When the temperature is large in comparison with the level splitting, $\beta \ll ml^2$, it is possible to approximate the sums (3.16) by integrals and so to evaluate the leading terms in expansions in powers of β/ml^2. The details of this

computation are given in appendix C. The results are

$$\Pi_0^{\mathrm{E}} = \frac{m}{\pi k^2}\left(\frac{1}{1+\mathrm{e}^{-\beta\mu}} + \frac{1}{96}\left(\frac{\beta}{ml^2}\right)^2 \sinh\frac{\beta\mu}{2}\operatorname{sech}^3\frac{\beta\mu}{2} + \ldots\right)$$

$$-\frac{\beta}{48\pi}\left(\operatorname{sech}^2\frac{\beta\mu}{2} - \frac{7}{120}\left(\frac{\beta}{ml^2}\right)^2\left\{\operatorname{sech}^2\frac{\beta\mu}{2} - \tfrac{3}{2}\operatorname{sech}^4\frac{\beta\mu}{2}\right\} + \ldots\right),$$

$$\Pi_1^{\mathrm{E}} = -\frac{1}{48\pi}\frac{\beta}{ml^2}\left(\operatorname{sech}^2\frac{\beta\mu}{2} - \frac{7}{120}\left(\frac{\beta}{ml^2}\right)^2\left\{\operatorname{sech}^2\frac{\beta\mu}{2} - \tfrac{3}{2}\operatorname{sech}^4\frac{\beta\mu}{2}\right\} + \ldots\right),$$

$$\Pi_2^{\mathrm{E}} = \frac{1}{12\pi m}\left(\frac{1}{1+\mathrm{e}^{-\beta\mu}} + \frac{7}{160}\left(\frac{\beta}{ml^2}\right)^2 \sinh\frac{\beta\mu}{2}\operatorname{sech}^3\frac{\beta\mu}{2} + \ldots\right), \tag{3.33}$$

where the chemical potential μ is determined by the formula

$$\nu = ml^2\mu + \frac{ml^2}{\beta}\ln(1+\mathrm{e}^{-\beta\mu}) - \frac{1}{96}\left(\frac{\beta}{ml^2}\right)\operatorname{sech}^2\frac{\beta\mu}{2} + \mathrm{O}\left(\frac{\beta}{ml^2}\right)^3. \tag{3.34}$$

The approximations are valid if $\beta/ml^2 \ll \beta\mu$. If we impose the stronger condition that $\beta\mu \gg 1$ then eq. (3.34) can be solved by iteration to give

$$\mu = \frac{\nu}{ml^2} - \frac{1}{\beta}\mathrm{e}^{-\beta\nu/ml^2} + \ldots. \tag{3.35}$$

This expression can be used to eliminate μ from eq. (3.33), leading to series expansions in powers of the small quantity $\exp(-\beta\nu/ml^2)$. The steps are justified if $\nu \gg 1$, in which case the high-temperature approximations (3.33) can be looked upon as expansions in powers of $1/\nu$.

4. Static couplings of electromagnetic currents

The euclidean action functional obtained in the last section can be used to evaluate the thermal averages of interaction energies associated with static current distributions. Here we shall consider only the simplest case, computation of the functional $W(I,J)$ with $J_\mu = 0$ and I_μ small. With the understanding that the background equations which determine the equilibrium are satisfied, we can treat I_μ as a weak perturbation and determine the response by solving the linear equations

$$\frac{\delta\Gamma^{\mathrm{E}}}{\delta A_\mu} = -I_\mu \quad \text{and} \quad \frac{\delta\Gamma^{\mathrm{E}}}{\delta a'_\mu} = 0, \tag{4.1}$$

where Γ^{E} comprises the bilinear parts of the classical and 1-loop terms

$$\Gamma^{E} = \int d^3x \left(\frac{1}{4e^2} F_{\mu\nu} F_{\mu\nu} + \frac{i\nu}{2\pi} \varepsilon_{\mu\lambda\rho} (a'_\mu \partial_\lambda a'_\rho) \right)$$

$$+ \tfrac{1}{2} \int d^3x\, d^3x'\, \Gamma^{E}_{\mu\nu}(x-x') \big(A_\mu(x) + a'_\mu(x) \big(A_\nu(x') + a'_\nu(x') \big) \big) . \quad (4.2)$$

On substituting the solution of the inhomogeneous equations (4.1) into (4.2) and taking the Legendre transform, as discussed in sect. 3, we obtain the functional

$$W(I,0) = \Gamma^{E}(A,a) + \int d^3x I_\mu A_\mu$$

$$= \tfrac{1}{2} \int d^3x I_\mu A_\mu ,$$

an expression which is bilinear in the external current distribution, I_μ,

$$W(I,0) = \tfrac{1}{2} \int d^3x\, d^3x'\, I_\mu(x) D_{\mu\nu}(x-x') I_\nu(x')$$

$$= \frac{1}{2\beta} \sum_s \int \frac{d^2k}{(2\pi)^2} \tilde{I}_\mu(-\omega_s, -k) \tilde{D}_{\mu\nu}(\omega_s, k) \tilde{I}_\nu(\omega_s, k) . \quad (4.3)$$

Since we are interested here only in static current distributions we shall take

$$\tilde{I}_\mu(\omega_s, k) = \delta_{s0} \beta \tilde{I}_\mu(k) ,$$

in which case (4.3) reduces to the form

$$W(I,0) = \frac{\beta}{2} \int \frac{d^2k}{(2\pi)^2} \tilde{I}_\mu(-k) \tilde{D}_{\mu\nu}(0,k) \tilde{I}_\nu(k)$$

$$= \beta F(I) , \quad (4.4)$$

where $F(I)$ represents the free energy of the given current distribution. The matrix $D_{\mu\nu}(x-x')$ can of course be interpreted as the thermal Green function for the electromagnetic field, and the functional $F(I)$ represents the interaction energy due to 1-photon exchange, including the effects due to mixing between electromagnetic and Chern–Simons fields. We shall show that this mixture simulates the exchange of two distinct massive states. One of these states may be interpreted as a massive "photon" and its presence is indicative of a kind of Meissner effect. The

other state, which is generally much heavier, probably plays the role of the Higgs state, familiar in Landau Gizburg theory, although here it is a vector rather than a scalar. In addition, the functional $F(I)$ will be found to contain a long-range static magnetic interaction $\sim -\tilde{I}_j \tilde{I}_j / k^2$ which vanishes at zero temperature.

The equations of motion (4.1) take the form

$$\frac{1}{e^2}\left(k^2 \tilde{A}_\mu - k_\mu k_\nu \tilde{A}_\nu\right) + \tilde{\Gamma}^{\mathrm{E}}_{\mu\nu}(k)\left(\tilde{A}_\nu + \tilde{a}'_\nu\right) = \tilde{I}_\mu ,$$

$$-\frac{\nu}{\pi}\varepsilon_{\mu\lambda\rho}k_\lambda \tilde{a}'_\rho + \tilde{\Gamma}^{\mathrm{E}}_{\mu\nu}(k)\left(\tilde{A}_\nu + \tilde{a}'_\nu\right) = 0 , \tag{4.5}$$

where the components of $\tilde{\Gamma}^{\mathrm{E}}_{\mu\nu}$ are given by eqs. (3.15) and (3.16). In addition to the static condition, $k_\tau = 0$, we shall impose the gauge conditions $k_\mu \tilde{A}_\mu = k_\mu \tilde{a}'_\mu = 0$. It is then a simple algebraic exercise to construct the solution

$$\tilde{A}_\mu = \tilde{D}_{\mu\nu}\tilde{I}_\nu .$$

The components of the propagator can be expressed in terms of three independent invariants,

$$D_{\tau\tau} = D_0 ,$$

$$D_{\tau j} = -D_{j\tau} = \varepsilon_{jl}k_l D_1 ,$$

$$D_{jl} = \left(\delta_{jl} - \frac{k_j k_l}{k^2}\right) D_2 . \tag{4.6}$$

The expressions for D_0, D_1 and D_2 are not very transparent and we shall simplify their appearance by introducing a number of parameters to characterize the low-momentum behaviour of the tensor $\tilde{\Gamma}^{\mathrm{E}}_{\mu\nu}$ defined in eq. (3.15)

$$\Pi^{\mathrm{E}}_0 = \frac{a}{k^2} + b, \qquad \Pi^{\mathrm{E}}_1 = c, \qquad \Pi^{\mathrm{E}}_2 = d . \tag{4.7}$$

These temperature-dependent parameters are defined by the sums (3.16). A lengthy but straightforward computation gives

$$D_0 = -\frac{1}{\Delta}\left[\frac{1}{e^2}\left(1 - \frac{\pi}{\nu}c\right)^2 + d\left(1 + \frac{1}{e^2}\left(\frac{\pi}{\nu}\right)^2 (a + k^2 b)\right)\right],$$

$$D_1 = -\frac{1}{2k^2\Delta}\left[c\left(1 - \frac{\pi}{\nu}c\right) - \frac{\pi}{\nu}(a + k^2 b)d\right],$$

$$D_2 = -\frac{1}{\Delta}\left[\frac{1}{e^2}\left(1 - \frac{\pi}{\nu}c\right)^2 + \left(\frac{a}{k^2} + b\right)\left(1 + \frac{1}{e^2}\left(\frac{\pi}{\nu}\right)^2 k^2 d\right)\right], \tag{4.8}$$

where Δ is quadratic in k^2

$$\Delta = \frac{a}{e^2} + c^2 + ad + \frac{k^2}{e^4}\left[\left(1 - \frac{\pi}{\nu}c\right)^2 + e^2 b + e^2 d + e^4 bd + \left(\frac{\pi}{\nu}\right)^2 ad\right] + \left(\frac{k^2}{e^4}\right)^2 \left(\frac{\pi}{\nu}\right)^2 e^4 bd . \tag{4.9}$$

To gain some understanding of these quantities it is essential to make approximations. We shall therefore assume, firstly, that we are in the low-temperature regime, $\beta \gg ml^2$, and, secondly, that the Chern–Simons parameter is large, $\nu \gg 1$. The formulae (3.32) then give

$$a = \frac{4m}{\pi} y\, e^{-y}, \qquad b = \frac{m}{n_e}\left(\frac{\nu}{\pi}\right)^2 (1 - 4y\, e^{-y}),$$

$$c = \frac{\nu}{\pi}(1 - 4y\, e^{-y}), \qquad d = \frac{\nu^2}{\pi m}(1 - 4y\, e^{-y}), \tag{4.10}$$

where $y = \beta/2ml^2$. Higher powers of e^{-y} and π/ν are neglected in eq. (4.10)*. Among the dimensionful parameters e^2, n_e and m there are two dimensionless ratios for which we shall adopt the estimates used by Hosotani and Chakravarty [8]. These authors regard the Maxwell lagrangian as a 3-dimensional remnant of the standard 4-dimensional theory, obtained by integrating over the thickness, $\delta \sim 10^{-7}$ cm, of some crystal layer. This suggests the value

$$e^2 \sim e_4^2/\delta \sim (10^{-5}\ \text{cm})^{-1}.$$

They also assume that the density of charge carriers is similar to the density of ions in the lattice

$$n_e \sim (10^{-7}\ \text{cm})^{-2}.$$

Finally, they assume that the "electron" mass is the usual one

$$m \sim 10^{10}\ \text{cm}^{-1}.$$

With these estimates we have

$$\frac{e^2}{m} \sim 10^{-5} \quad \text{and} \quad \frac{n_e}{m^2} \sim 10^{-6}, \tag{4.11}$$

and it becomes feasible to make significant simplifications.

* At zero temperature, $y \to \infty$, we have $a = 0$ while the coefficients b, c and d reduce to Π_0, Π_1 and $-\Pi_2$, respectively, given by eq. (2.15). Also in this limit the expression (4.9) reduces to (2.22) evaluated at $\omega = 0$.

Firstly, in order to justify the low-temperature expansions, we should require $y \geqslant 5$ (say). This implies

$$\frac{1}{\beta} = \frac{1}{2ml^2 y} = \frac{n_e}{2m}\frac{\pi}{\nu}\frac{1}{y} \lesssim \frac{\pi}{\nu} \times 10^3 \text{ K}. \tag{4.12}$$

Secondly, the dominant parts of Δ are given by

$$\Delta \simeq \frac{a}{e^2} + c^2 + k^2 e^{-2} b + \left(\frac{\pi}{\nu}\frac{k^2}{e^2}\right)^2 bd$$

$$= \left(\frac{\pi}{\nu e^2}\right)^2 bd\left(k^2 + M_+^2\right)\left(k^2 + M_-^2\right), \tag{4.13}$$

where the masses are given approximately by

$$M_+^2 = \frac{e^2}{b}\left(\frac{a}{e^2} + c^2\right)$$

$$= \frac{e^2 n_e}{m}(1 - 4y\,e^{-y}) + \frac{4\pi n_e}{\nu^2} y\,e^{-y},$$

$$M_-^2 = \left(\frac{\nu}{\pi}\right)^2 \frac{e^2}{d}$$

$$= \frac{e^2 m}{\pi}(1 + 4y\,e^{-y}). \tag{4.14}$$

Here we have discarded terms of order

$$\left(\frac{\pi}{\nu}\right)^2 \frac{d}{b}\left(\frac{a}{e^2} + c^2\right) \sim \frac{n_e}{m^2} \sim 10^{-6}$$

is comparison with unity. The value of M_+^2 at zero temperature coincides with the formula for penetration depth given by others [4, 8]. Our result gives the low-temperature corrections and we see that M_+^2 *increases* with temperature.

Although, regrettably, this model does not appear to undergo a phase transition at finite temperature*, we can define a pseudo-critical temperature at which the thermal contribution to the Meissner effect overtakes the zero-temperature contribution. The two contributions to M_+^2, according to the approximate formula (4.14),

* One should bear in mind that this is a perturbative result. It is conceivable that non-perturbative effects might lead to a qualitatively different conclusion, particularly in the case of small ν where our methods cannot be applied. (See note added in proof.)

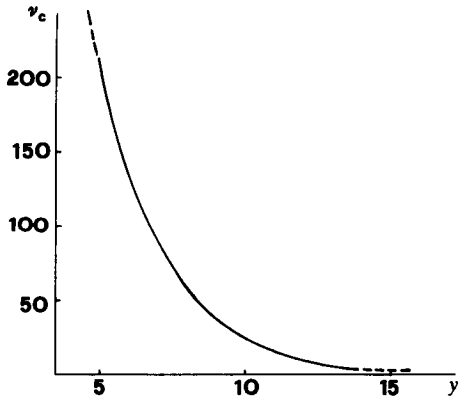


Fig. 2. Plot of pseudocritical coupling, $\nu_c(y)$. For $\nu < \nu_c$ the penetration depth is dominated by the thermal contribution. The curve is reliable provided ν and $y = 1/2ml^2T$ are not small.

become comparable at a temperature-dependent coupling, $\nu_c(y)$ given by

$$\nu_c^2 = \frac{4\pi m}{e} y\, e^{-y}(1 + 4y\, e^{-y}). \tag{4.15}$$

The expression for M_+^2 then takes the form

$$M_+^2 = \frac{e^2 n_e}{m}\left(1 + \frac{\nu_c^2}{\nu^2}\right)(1 - 4y\, e^{-y}), \tag{4.16}$$

and we see that the thermal contribution dominates for $\nu < \nu_c$. The pseudocritical coupling ν_c is plotted in fig. 2 for a range of temperatures.

The invariant components (4.8) reduce, in the low-temperature regime and with the neglect of higher powers of the ratios (4.11), to the following form

$$\begin{aligned}
-D_0 &= \frac{e^2}{k^2 + M_-^2} + \frac{e^2}{k^2 + M_+^2}\frac{8\pi n_e}{m^2} y\, e^{-y}, \\
-D_1 &= -\frac{1}{k^2 + M_+^2}\frac{\nu e^4 n_e}{2M_+^2 m^2}\left[1 - 4y\, e^{-y}\left(2 - \frac{\pi m}{\nu^2 e^2}\right)\right] \\
&\quad + \frac{1}{k^2 + M_-^2}\frac{\nu e^4}{2\pi M_-^2}\left[1 + 4y\, e^{-y}\left(\frac{\pi}{\nu}\right)^2 \frac{n_e}{me^2}\right], \\
-D_2 &= \frac{e^2}{k^2 + M_+^2}\left[1 - \frac{4\pi m}{\nu^2 e^2} y\, e^{-y}(1 + 8y\, e^{-y})\left(1 + \frac{4\pi m}{\nu^2 e^2} y\, e^{-y}\right)^{-1}\right] \\
&\quad + \frac{m}{k^2}\left(\frac{4\pi}{\nu^2}\right) y\, e^{-y}(1 + 8y\, e^{-y})\left(1 + \frac{4\pi m}{\nu^2 e^2} y\, e^{-y}\right)^{-1}.
\end{aligned} \tag{4.17}$$

In order to have a rough idea of the relative strengths of the various forces involved here, we give the energy functional at zero temperature with the

Chern–Simons parameter assigned to an intermediate range such that

$$\frac{n_e}{me^2} \ll \left(\frac{\nu}{\pi}\right)^2 \ll \frac{m}{e^2}, \tag{4.18}$$

which means $1 \ll \nu \ll 10^3$. At the same time we restore the real charge density $I_0 = iI_\tau$. The result is

$$E(I) = \tfrac{1}{2}\int \frac{d^2k}{(2\pi)^2}\left\{\tilde{I}_0(-k)\frac{e^2}{k^2+M_-^2}\tilde{I}_0(k)\right.$$
$$+\tilde{I}_0(-k)\left[\frac{1}{k^2+M_-^2}-\frac{1}{k^2+M_+^2}\right]\frac{\nu e^2}{m}\varepsilon_{jl}ik_l\tilde{I}_j(k)$$
$$\left.-\tilde{I}_j(-k)\frac{e^2}{k^2+M_+^2}\tilde{I}_j(k)\right\}. \tag{4.19}$$

This indicates that the static force between currents is of intermediate range and attractive. There is a two-component force, one of which is longer ranged, between charge and current.

To complete the discussion, we now consider the high temperature regime, $\beta \ll ml^2$. If $\nu \gg 1$ then we have, according to eq. (3.35)

$$\mu \simeq \frac{\nu}{ml^2} = \pi\frac{n_e}{m} \tag{4.20}$$

and the limiting forms of the expressions (3.33) give

$$a \simeq \frac{m}{\pi}, \qquad b \simeq 0 \simeq c, \qquad d \simeq \frac{1}{12\pi m}, \tag{4.21}$$

which can be substituted into eqs. (4.8) and (4.9). We obtain

$$D_0 \simeq -\frac{e^2}{k^2+me^2/\pi},$$
$$D_1 \simeq \frac{1}{24}\frac{e^2}{\nu m}\left(\frac{1}{k^2}-\frac{1}{k^2+me^2/\pi}\right),$$
$$D_2 \simeq -\frac{e^2}{k^2}. \tag{4.22}$$

The pole at $k^2 = 0$ in D_2 indicates the presence of a long-range magnetic interaction. This does not imply a phase transition, however. Indeed, the long-range effect, which derives from the pole in Π_0^E, is present at any finite temperature. Its

coupling strength depends on the temperature and, as can be seen from the low-temperature expression for D_2 given by eq. (4.17), this coupling goes smoothly to zero at zero temperature. The pure Meissner effect which is observed at zero temperature, for example in the formula (4.19), is certainly compromised by the appearance of this pole in the finite temperature amplitude. We must conclude that the magnetic field does in fact develop a long-range component along with the finite range ($\sim M_+^{-1}$) component which dominates at low temperature. What we have found is that the relative strengths of these components varies smoothly with temperature and that the long-range part comes to dominate at high temperatures.

5. Conclusions

In this paper we have applied standard field-theoretic perturbation theory to calculate thermal corrections to the Meissner effect recently demonstrated in a simple model of Chern–Simons superconductivity [4, 8]. The calculation is expressed in the language of effective action functionals – minkowskian for the zero temperature case and euclidean for the finite temperature case. Our main approximations are the restriction to weak and slowly varying fields, and the neglect of 2-loop contributions. The model involves two coupling parameters, e^2, with the dimensions of mass, and π/ν, which is dimensionless. Both of these parameters must be small in order to justify the 1-loop approximation. It may eventually become possible to improve on this, extending the range to small values of the Chern–Simons parameter, ν, by applying the methods of fractional statistics, but we have not attempted such a calculation. In this paper the $(2+1)$-dimensional "electrons" are treated as ordinary fermions.

The mechanism which gives rise to a kind of Meissner effect in this model is the induction, at 1-loop level, of a Chern–Simons term for the electromagnetic potential. As we have pointed out in sect. 1, the presence of such a term along with the standard Maxwell kinetic term leads to a massive photon. Since there are two independent gauge fields in this model it is necessary to take account of their mixing, which is substantial. There are in fact two massive states representing the independent mixtures. Because of the assumed values of the parameters, n_e, m and e^2, one of the states is considerably lighter than the other and it is reasonable to consider it as the massive "photon". However, it should be kept in mind that there are two independent states which mediate the interaction between currents. The Meissner effect in this model is kinematically more complicated than in the classic Ginzburg–Landau description. At finite temperatures, the situation is further complicated by the appearance of a long-range component in the static interactions. We have shown that the relative strengths of the long and finite range components varies continuously with temperature.

It is a great virtue of this model that the electron energies are quantized. The level splitting is inversely proportional to the Chern–Simons parameter, ν. Rela-

tive to this splitting it is possible to define low- and high-temperature regimes, where the sums which define the thermal averages can be evaluated approximately. For the values assumed in this paper, the level splitting is given by $(ml^2)^{-1} = \pi n_e/m\nu \sim 10^4$ K$/\nu$ and this means that the physically interesting temperatures, $\sim 10^2$ K, may lie in either the low or high regime, depending on the value of ν. If ν is very large then we can view the high-temperature approximation as an expansion in powers of $1/\nu$.

It was, of course, a disappointment to find no indication of a conventional phase transition in the temperature dependence of the penetration depth. However, this conclusion seems to be inescapable. The usual criterion for the existence of a critical temperature, manifested as a divergence in the penetration depth, would be the vanishing at $k^2 = 0$ of the determinant, $\Delta(k^2)$, given by eq. (4.9). But the value

$$\Delta(0) = \frac{a}{e^2} + c^2 + ad,$$

where a, c and d are defined by eqs. (4.7) and (3.16), seems to be inevitably positive. The coefficient a is given as a sum of positive terms and e^2d is relatively small. In the high- and low-temperature approximations we have shown that e^2d is in fact positive. Although we have not been able to show that this is always the case it is hard to see how it could become sufficiently large and negative to make $\Delta(0)$ vanish. However, our conclusion is based on the 1-loop approximation which makes sense only if π/ν is small. A more sophisticated approach would be necessary if ν is small, and the result could well be different*.

The Chern–Simons parameter must be an integer. This restriction on the allowed values of ν is based on two independent considerations. Firstly, as discussed in sect. 2, the consistency of the assumed ground-state configuration, with a uniform Chern–Simons magnetic background, requires that $|\nu|$ should equal the number of filled Landau levels and therefore be a positive integer [4, 8]. Secondly, the lagrangian density (2.1) is invariant with respect to gauge transformation only up to a total derivative. If the theory is formulated on a compact space then it becomes possible to consider gauge transformations which are topologically non-trivial and for which the integral of the total derivative does not vanish. Invariance of the phase, e^{iS}, with respect to such transformations is ensured only if ν is an integer [2, 8]. On the other hand, we should point out that the temperature-dependent radiative correction to ν, represented by the coefficient c, are not integral. This seems to indicate a breakdown of the topologically non-trivial parts of the gauge symmetry.

* For example, such a difference might come about because of higher order effects. These could be estimated by a renormalization group approach which would make the coupling parameters, e^2 and ν, into functions of the temperature.

It would be interesting to see if any modifications of the present simple model are able to exhibit critical behaviour at finite temperature. Even more interesting would be the discovery of some explanation, in terms of a more fundamental microscopic theory, for the appearance of a Chern–Simons field and its associated parameter, ν.

Note added in proof

After submitting this work we found that our discussion of the high-temperature regime was incomplete. We overlooked the fact that the parameter $\Pi_1^E = c$ changes sign (compare (3.32) with (C.24)) at a finite temperature. This implies that M_-^2 goes to infinity at this temperature and becomes formally negative at higher temperatures. The consequent instability of our translation invariant ground state indicates the breakdown of the perturbation scheme and the appearance of a new phase.

We are indebted to E. Witten for a very useful discussion which stimulated our interest in this subject. We are also grateful to Y. Hosotani for pointing out an error in an early version of this paper.

Appendix A

KINEMATICS OF LANDAU STATES

Because of the presence of a uniform magnetic field, the 1-electron hamiltonian is not invariant with respect to rotations and translations. However, because the field is uniform it should be possible to compensate for the action of these space transformations by means of gauge transformations. In this appendix we shall construct the appropriate gauge action and show that the hamiltonian is indeed invariant with respect to the combined group.

The background Chern–Simons field, used in sect. 2, is given by

$$a_1(x) = -x_2, \qquad a_2(x) = 0. \tag{A.1}$$

Here, in order to streamline the notation we shall adopt a system of units in which the magnetic length, l, is equal to unity. The action on the Chern–Simons vector of an arbitrary space transformation

$$x_1 \to x_1' = x_1 \cos\theta + x_2 \sin\theta + \varepsilon_1,$$
$$x_2 \to x_2' = -x_1 \sin\theta + x_2 \cos\theta + \varepsilon_2,$$

followed by a gauge transformation, is given by

$$a_1(x) \to a_1'(x') = a_1(x)\cos\theta + a_2(x)\sin\theta + \partial_1 \Lambda,$$
$$a_2(x) \to a_2'(x') = -a_1(x)\sin\theta + a_2(x)\cos\theta + \partial_2 \Lambda. \tag{A.2}$$

In order to have $a_i'(x) = a_i(x)$ or, more specifically,

$$a_1'(x') = -x_2', \qquad a_2'(x') = 0,$$

for arbitrary θ and ε_j, it is necessary to choose

$$\Lambda(x) = \tfrac{1}{2}(x_1^2 - x_2^2)\sin\theta - x_1\varepsilon_2 . \tag{A.3}$$

This is the compensating gauge transformation. The action of the combined transformation on the electron field is given by*

$$\psi(x) \to \psi'(x') = e^{i\Lambda(x)}\psi(x) . \tag{A.4}$$

We can define the generators of infinitesimal transformations in the 1-electron Hilbert space by expanding eq. (A.4)

$$\begin{aligned}\delta\psi(x) &= \psi'(x) - \psi(x) \\ &= -i\varepsilon_j P_j\psi(x) + i\theta J\psi(x) .\end{aligned}$$

This gives

$$\begin{aligned} P_1\psi &= -i\partial_1\psi , \\ P_2\psi &= (-i\partial_2 + x_1)\psi , \\ J\psi &= -i(x_1\partial_2 - x_2\partial_1)\psi + \tfrac{1}{2}(x_1^2 - x_2^2)\psi . \end{aligned} \tag{A.5}$$

It is, of course, true that the space group does not commute with the gauge group and, in deriving eq. (A.5) we have required the compensating gauge transformation to follow the space transformation. The non-commutativity of these groups gives rise to a modification of the Lie algebra of the effective invariance group in comparison with that of the pure space group. Indeed, from eq. (A.5) it follows that

$$[P_1, J] = -iP_2, \qquad [P_2, J] = iP_1, \qquad [P_1, P_2] = -i . \tag{A.6}$$

Notice that the "translation" generators do not commute. Instead, they behave like canonically conjugate variables.

It is instructive and useful to replace the original canonical variables x_j and $\pi_j = -i\partial_j$ by the liner combinations,

$$\begin{aligned} Q &= \pi_1 + x_2, \qquad & q &= \pi_2 + x_1 , \\ P &= \pi_2 , & p &= \pi_1 , \end{aligned} \tag{A.7}$$

* We treat the electron field as a space scalar.

which define two independent canonical pairs (q, p) and (Q, P). It is a simple exercise to express the generators (A.5) in terms of the new variables

$$P_1 = p, \qquad P_2 = q,$$

$$J = \tfrac{1}{2}(q^2 + p^2) - \tfrac{1}{2}(Q^2 + P^2). \tag{A.8}$$

Moreover, the 1-electron hamiltonian becomes

$$H = \frac{1}{2m}(Q^2 + P^2), \tag{A.9}$$

which clearly commutes with the generators (A.8).

Since the angular momentum is now expressed as the difference of two independent oscillator hamiltonians (one the true hamiltonian and the other a kind of shadow hamiltonian that makes no contribution to the energy) it will be relatively easy to perform rotation invariant computations by using a basis of oscillator eigenstates

$$|n, r\rangle = \frac{1}{\sqrt{n!}}\left(\frac{Q - iP}{\sqrt{2}}\right)^n \frac{1}{\sqrt{r!}}\left(\frac{q - ip}{\sqrt{2}}\right)^r |0\rangle, \tag{A.10}$$

where the ground state is defined by

$$(Q + iP)|0\rangle = (q + ip)|0\rangle = 0. \tag{A.11}$$

With respect to the coordinate basis $|Q, q\rangle$ the wave functions factorize

$$\langle Q, q | n, r\rangle = v_n(Q) v_r(q), \tag{A.12}$$

where the v's are expressed in the familiar way in terms of Hermite polynomials

$$\langle Q | n\rangle = v_n(Q) = (\sqrt{\pi}\, 2^n n!)^{-1/2} H_n(Q) e^{-Q^2/2}. \tag{A.13}$$

We shall have occasion to employ a momentum basis $|P\rangle$ as well. These states are normalized according to the convention

$$\langle Q | Q'\rangle = \delta(Q - Q'),$$

$$\langle P | P'\rangle = 2\pi\delta(P - P') \tag{A.14}$$

and the transition amplitudes are given by

$$\langle Q | P\rangle = e^{iQP} \tag{A.15}$$

Appendix B

THE ZERO-TEMPERATURE LOOP INTEGRAL

The 1-electron Green function of sect. 2 satisfies the differential equation

$$(i\partial_0 - H)G(x,x') = -\delta_3(x-x')$$

and it can be expressed in the form of an integral over frequencies

$$G(x,x') = \int_{C_N} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \langle x|G(\omega)|x'\rangle , \tag{B.1}$$

where the operator $G(\omega)$ is defined by

$$G(\omega) = (H-\omega)^{-1}, \tag{B.2}$$

provided ω does not coincide with any of the eigenvalues of H. The arrangement of the integration contour C_N is described in sect. 2. Here we shall be concerned mainly with the matrix elements of $G(\omega)$. We shall also need expressions for the derivatives $\nabla_i G(x,x')$, $\nabla_i' G(x,x')$, etc. In the notation of appendix A we have

$$\nabla_1 = \partial_1 + ix_2 = i(\pi_1 + x_2) = iQ ,$$

$$\nabla_2 = \partial_2 = i\pi_2 = iP \tag{B.3}$$

and it is useful to define the 2-vector Π_i according to

$$\Pi_1 = Q, \qquad \Pi_2 = P. \tag{B.4}$$

In terms of this operator the covariant derivatives of ψ and ψ^+ are expressed by

$$\nabla_i \psi = i\Pi_i \psi \quad \text{and} \quad \nabla_i \psi^+ = -i\psi^+ \Pi_i . \tag{B.5}$$

It follows that the covariant derivatives $\nabla_i G(x,x')$ and $\nabla_i' G(x,x')$ can be represented by contour integrals like (B.1) with the operator $G(\omega)$ replaced by $i\Pi_i G(\omega)$ and $-iG(\omega)\Pi_i$, respectively. Higher order covariant derivatives of $G(x,x')$ are given by obvious generalizations of this rule.

Let us now consider the structure of the 1-loop ground state polarization tensor. In the typical 1-loop computation we arrive at the integral

$$\int d^2x\, d^2x'\, e^{-ik\cdot x + ik'\cdot x'} \langle x|G(\omega_1)|x'\rangle\langle x'|G(\omega_2)|x\rangle = \mathrm{Tr}\left(e^{-ik\cdot x} G(\omega_1) e^{-ik'\cdot x'} G(\omega_2)\right). \tag{B.6}$$

However, because the operator $G(\omega)$ involves only the canonical pair (Q, P) and not the pair (q, p) we can affect an immediate simplification. Substituting from eq. (A.7)

$$\begin{aligned} \boldsymbol{k}\cdot\boldsymbol{x} &= k_1(q-P)+k_2(Q-p) \\ &= k_1 q - k_2 p - k_1 P + k_2 Q \\ &= k_1 q - k_2 p + \boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}, \end{aligned} \tag{B.7}$$

where $\tilde{\Pi}_i$ denotes the dual of Π_i defined above

$$\tilde{\Pi}_1 = -\Pi_2, \quad \tilde{\Pi}_2 = \Pi_1. \tag{B.8}$$

The trace (B.6) factorizes

$$\begin{aligned} &\mathrm{Tr}\left(\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} G(\omega_1)\mathrm{e}^{i\boldsymbol{k}'\cdot\boldsymbol{x}} G(\omega_2)\right) \\ &\quad = \mathrm{Tr}\left(\mathrm{e}^{-ik_1 q + ik_2 p}\,\mathrm{e}^{ik_1' q - ik_2' p}\right)\mathrm{Tr}\left(\mathrm{e}^{-i\boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}} G(\omega_1)\mathrm{e}^{i\boldsymbol{k}'\cdot\tilde{\boldsymbol{\Pi}}} G(\omega_2)\right). \end{aligned}$$

The first factor can be evaluated explicitly

$$\begin{aligned} \mathrm{Tr}\left(\mathrm{e}^{-ik_1 q + ik_2 p}\,\mathrm{e}^{ik_1' q - ik_2' p}\right) &= \mathrm{Tr}\left(\mathrm{e}^{-i(k_1 - k_1')q}\,\mathrm{e}^{i(k_2 - k_2')p}\right)\mathrm{e}^{i(k_1 k_2 - k_1' k_2')/2} \\ &= \int \mathrm{d}q\,\mathrm{e}^{-i(k_1 - k_1')q}\left\langle q\right|\mathrm{e}^{i(k_2 - k_2')p}\left|q\right\rangle \mathrm{e}^{i(k_1 k_2 - k_1' k_2')/2} \\ &= 2\pi\delta_2(\boldsymbol{k}-\boldsymbol{k}'). \end{aligned} \tag{B.9}$$

Thus, the integral (B.6) reduces to the form

$$\begin{aligned} &\mathrm{Tr}\left(\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} G(\omega_1)\mathrm{e}^{i\boldsymbol{k}'\cdot\boldsymbol{x}} G(\omega_2)\right) \\ &\quad = 2\pi\delta_2(\boldsymbol{k}-\boldsymbol{k}')\mathrm{Tr}\left(\mathrm{e}^{-i\boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}} G(\omega_1)\mathrm{e}^{i\boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}} G(\omega_2)\right) \\ &\quad = 2\pi\delta_2(\boldsymbol{k}-\boldsymbol{k}')\sum_{n,m}\frac{\left\langle n\right|\mathrm{e}^{-i\boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}}\left|m\right\rangle\left\langle m\right|\mathrm{e}^{i\boldsymbol{k}\cdot\tilde{\boldsymbol{\Pi}}}\left|n\right\rangle}{(\varepsilon_m-\omega_1)(\varepsilon_n-\omega_2)}. \end{aligned} \tag{B.10}$$

Generalization to integrals which involve covariant derivatives of the Green function is immediate. One has only to make the appropriate modifications, $G(\omega_1) \to i\Pi_i G(\omega_1)$, etc. The rotation and translation invariance of the polarization tensor is manifest in these expressions.

The oscillator matrix elements in eq. (B.10) can be evaluated to any desired accuracy by employing the wave functions (A.13). In this paper we are interested

only in the leading terms in an expansion in powers of k. For this it is sufficient to use algebraic methods. To illustrate, we define the irreducible components

$$k_{\pm} = \frac{1}{\sqrt{2}}(k_1 \pm ik_2) \tag{B.11}$$

and write

$$\begin{aligned} ik \cdot \tilde{\Pi} &= -ik_1 P + ik_2 Q \\ &= k_+ a^+ - k_- a, \end{aligned} \tag{B.12}$$

where a^+ and a are the raising and lowering operators

$$a = \frac{1}{\sqrt{2}}(Q + iP), \qquad a^+ = \frac{1}{\sqrt{2}}(Q - iP). \tag{B.13}$$

The oscillator eigenstates are defined by

$$|n\rangle = \frac{1}{\sqrt{n!}}(a^+)^n |0\rangle \tag{B.14}$$

and it is straightforward to evaluate the first few terms of the expansion

$$\begin{aligned} e^{ik\cdot\tilde{\Pi}}|n\rangle &= \sum_r \frac{1}{r!}(k_+ a^+ - k_- a)^r |n\rangle \\ &= |n\rangle + k_+\sqrt{n+1}\,|n+1\rangle - k_-\sqrt{n}\,|n-1\rangle + \tfrac{1}{2}k_+^2\sqrt{(n+1)(n+2)}\,|n+2\rangle \\ &\quad - k_+k_-(n+\tfrac{1}{2})|n\rangle + \tfrac{1}{2}k_-^2\sqrt{n(n-1)}\,|n-2\rangle + \ldots, \end{aligned} \tag{B.15}$$

etc. It is clear that the matrix element $\langle m|e^{ik\Pi}|n\rangle$ has the threshold behaviour $k^{|m-n|}$ and it follows that, in evaluating the sum (B.10) to order k^{2s}, we need to keep only the terms with $|m-n| \leqslant s$.

It remains to consider the integration over frequencies. The most direct way to deal with these integrals is to displace the eigenvalues of H by infinitesimal imaginary parts and then take the contour C_N along the real axis. Specifically, we make the displacements

$$\begin{aligned} &\varepsilon_n \to \varepsilon_n + i\delta, \qquad 0 \leqslant n < N, \\ &\varepsilon_n \to \varepsilon_n - i\delta, \qquad n \geqslant N, \end{aligned} \tag{B.16}$$

where δ is real and positive. The computation then proceeds in a straightforward

way. To illustrate, we consider the polarization component Γ_{00} defined in eq. (2.12). The 3-dimensional Fourier transform is given by

$$\begin{aligned}\tilde{\Gamma}_{00}(k,k') &= \int \mathrm{d}^3x\,\mathrm{d}^3x'\,\mathrm{e}^{-ikx+ik'x'}\,i\,\mathrm{Tr}\,G(x,x')G(x',x)\\ &= 2i\int \mathrm{d}t\,\mathrm{d}t'\,\mathrm{e}^{i\omega t-i\omega' t'}\int\frac{\mathrm{d}\omega_1}{2\pi}\mathrm{e}^{-i\omega_1(t-t')}\\ &\quad\times\int\frac{\mathrm{d}\omega_2}{2\pi}\mathrm{e}^{+i\omega_2(t-t')}\,\mathrm{Tr}\big(\mathrm{e}^{-ik\cdot x}G(\omega_1)\mathrm{e}^{ik\cdot x}G(\omega_2)\big)\\ &= (2i)2\pi\delta(\omega-\omega')\int\frac{\mathrm{d}\omega_1}{2\pi}\mathrm{Tr}\big(\mathrm{e}^{-ik\cdot x}G(\omega_1)\mathrm{e}^{ik\cdot x}G(\omega_1-\omega)\big)\\ &= 2\pi\delta(\omega-\omega')(2\pi)^2\delta(k-k')\tilde{\Gamma}_{00}(\omega,k)\,.\end{aligned}$$

In the first line of this development there is a trace over the spin degree of freedom which yields the factor 2 in the succeeding lines. Substituting the expression (B.10) and extracting the momentum conservation delta functions one obtains,

$$\begin{aligned}\tilde{\Gamma}_{00}(\omega,k) &= \frac{i}{\pi}\int\frac{\mathrm{d}\omega_1}{2\pi}\,\mathrm{Tr}\big(\mathrm{e}^{-ik\cdot\hat{\Pi}}G(\omega_1)\mathrm{e}^{ik\cdot\hat{\Pi}}G(\omega_1-\omega)\big)\\ &= \frac{i}{\pi}\sum_{m,n}\langle n|\mathrm{e}^{-ik\cdot\hat{\Pi}}|m\rangle\langle m|\mathrm{e}^{ik\cdot\hat{\Pi}}|n\rangle\\ &\quad\times\int\frac{\mathrm{d}\omega_1}{2\pi}(\varepsilon_m-\omega_1)^{-1}(\varepsilon_n-\omega_1+\omega)^{-1}.\end{aligned}\tag{B.17}$$

With the energies displaced off the real axis according to the rule (B.16) the integral over ω_1 is well defined and easy to evaluate. It is non-vanishing only if the two poles are situated on opposite sides of the real axis. Using Cauchy's theorem

$$\int\frac{\mathrm{d}\omega_1}{2\pi i}(\varepsilon_m-\omega_1)^{-1}(\varepsilon_n-\omega_1+\omega)^{-1}=\begin{cases}(\varepsilon_n-\varepsilon_m+\omega)^{-1}, & m\geqslant N,\ n<N,\\ -(\varepsilon_n-\varepsilon_m+\omega)^{-1}, & m<N,\ n\geqslant N,\\ 0, & \text{otherwise}.\end{cases}\tag{B.18}$$

Hence the result

$$\tilde{\Gamma}_{00}(\omega,k)=\frac{1}{\pi}\Big[-\sum_{m\geqslant N}\sum_{n<N}+\sum_{m<N}\sum_{n\geqslant N}\Big]\frac{\langle n|\mathrm{e}^{-ik\cdot\hat{\Pi}}|m\rangle\langle m|\mathrm{e}^{ik\cdot\hat{\Pi}}|n\rangle}{\varepsilon_n-\varepsilon_m+\omega}\,.\tag{B.19}$$

The other components of the polarization tensor are obtained in the same way

$$\tilde{\Gamma}_{j0}(\omega,k) = -\frac{1}{2m\pi}\left[-\sum_{m\geq N}\sum_{n<N}+\sum_{m<N}\sum_{n\geq N}\right]\frac{\langle n|\{e^{-ik\cdot\hat{\Pi}},\Pi_j\}|m\rangle\langle m|e^{ik\cdot\hat{\Pi}}|n\rangle}{\varepsilon_n-\varepsilon_m+\omega},$$

$$\tilde{\Gamma}_{jl}(\omega,k) = \frac{1}{4\pi m^2}\left[-\sum_{m\geq N}\sum_{n<N}+\sum_{m<N}\sum_{n\geq N}\right]\frac{\langle n|\{e^{-ik\cdot\hat{\Pi}},\Pi_j\}|m\rangle\langle m|\{e^{ik\cdot\hat{\Pi}},\Pi_l\}|n\rangle}{\varepsilon_n-\varepsilon_m+\omega}$$

$$-\frac{N}{\pi m}\delta_{jl},$$

where the symbol { , } denotes the anticommutator. On expanding in powers of k_j and ω, using the formula (B.15) for the oscillator matrix elements, the results (2.15) are obtained.

Appendix C

FINITE TEMPERATURE COMPUTATIONS

The 1-electron thermal Green function of sect. 3 satisfies the differential equation

$$(\partial_\tau + H - \mu)\mathcal{G}(x,x') = -\delta_3(x-x')$$

and it can be expanded in the form of a sum over discrete frequencies

$$\mathcal{G}(x,x') = -\frac{1}{\beta}\sum_s e^{-i\zeta_s(\tau-\tau')}\langle x|G(i\zeta_s+\mu)|x'\rangle, \qquad \text{(C.1)}$$

where $G(\omega)$ is the operator (B.2) defined in appendix B and

$$\zeta_s = \left(s+\frac{1}{2}\right)\frac{2\pi}{\beta}, \qquad s\in\mathbb{Z}.$$

The chemical potential, μ, is determined by the equilibrium conditions (3.17) as discussed in the main text. This determination depends on the formula (3.12) for the tadpole amplitude $\Gamma^{\mathrm{E}}_\mu(x)$ which we now derive. The component $\Gamma^{\mathrm{E}}_\tau(x)$ is defined by the first of equations (3.10)

$$\Gamma^{\mathrm{E}}_\tau(x) = i\,\mathrm{Tr}\,\mathcal{G}(x,x')\big|_{x'=x,\tau'=\tau+\delta}$$

$$= -\frac{2i}{\beta}\sum_s e^{i\zeta_s\delta}\langle x|G(i\zeta_s+\mu)|x\rangle, \qquad \text{(C.2)}$$

where δ is an infinitesimal positive quantity. For fixed $x = \hat{x}$, the diagonal matrix element in eq. (C.2) can be expressed as a trace in the 1-electron Hilbert space

$$\langle \hat{x}|G(i\zeta_s+\mu)|\hat{x}\rangle = \mathrm{Tr}(\delta(x-\hat{x})G(i\zeta_s+\mu))$$

$$= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \mathrm{e}^{-ik\cdot\hat{x}} \mathrm{Tr}(\mathrm{e}^{ik\cdot x} G(i\zeta_s+\mu))$$

$$= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \mathrm{e}^{ik\cdot\hat{x}} \mathrm{Tr}(\mathrm{e}^{ik_1q-ik_2p}) \mathrm{Tr}(\mathrm{e}^{-ik\cdot\bar{\Pi}} G(i\zeta_s+\mu)),$$

where we have used eq. (B.7) to separate the trace into independent factors. The first factor is purely kinematical

$$\mathrm{Tr}(\mathrm{e}^{ik_1q-ik_2p}) = 2\pi\delta_2(k),$$

according to eq. (B.9). Hence, we have

$$\langle \hat{x}|G(i\zeta_s+\mu)|\hat{x}\rangle = \frac{1}{2\pi} \mathrm{Tr}(G(i\zeta_s+\mu))$$

$$= \frac{1}{2\pi} \sum_n (\varepsilon_n - i\zeta_s - \mu)^{-1}, \tag{C.3}$$

and, on substituting into eq. (C.2),

$$\Gamma_\tau^{\mathrm{E}}(x) = \frac{i}{\pi} \sum_n \frac{1}{\beta} \sum_s (i\zeta_s - \varepsilon_n + \mu)^{-1} \mathrm{e}^{i\zeta_s\delta}. \tag{C.4}$$

The sum over s does not converge very well but it can be defined in a plausible way. Thus, by taking a derivative with respect to μ the convergence is improved and we can set $\delta = 0$. The resulting sum can be evaluated by contour methods,

$$-\frac{1}{\beta} \sum_s (i\zeta_s - \varepsilon_n + \mu)^{-2} = -\frac{i}{2\beta} \int_{\mathrm{C}} \frac{\mathrm{d}z}{\cot \pi z} \left(\frac{2\pi i}{\beta} z - \varepsilon_n + \mu \right)^{-2},$$

where the contour C encircles all the zeros of $\cot \pi z$, i.e. $z = s + \frac{1}{2}$, $s \in \mathbb{Z}$, in the positive sense. This contour can be replaced by a pair of straight lines, one running to the right just below the real axis and one running to the left just above the real axis. These two segments can be closed by means of large semicircles in the lower and upper half planes, respectively (see fig. 3). Their evaluation is then trivial since one or other of these contours will contain the dipole at $z = (\beta/2\pi i)(\varepsilon_n - \mu)$,

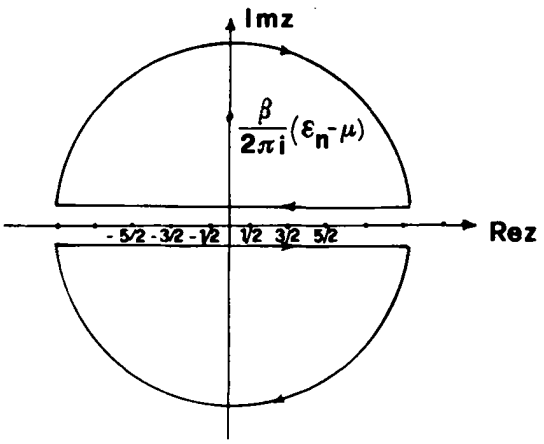


Fig. 3. The contour used in evaluating thermal sums. Poles on the real axis are given by the zeros of $\cot \pi z$ (appendix C).

depending on whether $\varepsilon_n - \mu$ is positive or negative. The result is

$$-\frac{1}{\beta}\sum_s (i\zeta_s - \varepsilon_n + \mu)^{-2} = \frac{\beta}{4}\operatorname{sech}^2\frac{\beta}{2}(\varepsilon_n - \mu). \tag{C.5}$$

This can be integrated to give, up to a constant,

$$\frac{1}{\beta}\sum_s (i\zeta_s - \varepsilon_n + \mu)^{-1} = \tfrac{1}{2}\left(1 - \operatorname{th}\frac{\beta}{2}(\varepsilon_n - \mu)\right)$$

$$= (e^{\beta(\varepsilon_n-\mu)} + 1)^{-1}. \tag{C.6}$$

The constant of integration is fixed by requiring appropriate behaviour in the zero-temperature limit, $\beta \to \infty$. Thus, we obtain the formula,

$$\Gamma_\tau^{\mathrm{E}}(x) = \frac{i}{\pi}\sum_n (e^{\beta(\varepsilon_n-\mu)} + 1)^{-1}. \tag{C.7}$$

It is easy to see that the space components Γ_j^{E} must vanish since, according to eq. (3.10), they must contain the factor $\mathrm{Tr}(\Pi_j G)$ which clearly vanishes.

Now consider the components of the polarization tensor

$$\Gamma_{\tau\tau}^{\mathrm{E}}(x,x') = -\mathrm{Tr}\,\mathscr{G}(x,x')\mathscr{G}(x',x)$$

$$= -\frac{2}{\beta^2}\sum_{s_1,s_2} e^{-i(\zeta_{s_1}-\zeta_{s_2})(\tau-\tau')}\langle x|G(i\zeta_{s_1}+\mu)|x'\rangle\langle x'|G(i\zeta_{s_2}+\mu)|x\rangle,$$

where we have substituted the expansion (C.1) and taken the trace over the spin degree of freedom. As in appendix B we take the Fourier transform and remove the delta functions associated with energy and momentum conservation to obtain

$$\tilde{\Gamma}^{\mathrm{E}}_{\tau\tau}(\omega_s, \boldsymbol{k}) = -\frac{1}{\pi\beta}\sum_{s_1} \mathrm{Tr}\left[\mathrm{e}^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{\Pi}}} G(i\zeta_{s_1}+\mu)\mathrm{e}^{i\boldsymbol{k}\cdot\hat{\boldsymbol{\Pi}}} G(i\zeta_{s_1}-i\omega_s+\mu)\right], \quad \text{(C.8)}$$

where $\omega_s = 2\pi s/\beta$, $s \in \mathbb{Z}$. The trace refers to the Hilbert space of the oscillator variables P and Q. (This formula could have been obtained directly from the zero temperature expression (B.17) by making the replacements $\omega \to i\omega_s$, $\omega_1 \to i\zeta_1 + \mu$ and

$$\int \frac{\mathrm{d}\omega_1}{2\pi} \to \frac{i}{\beta}\sum_{s_1}\Bigg).$$

On expressing the trace in eq. (C.8) in terms of oscillator matrix elements we arrive at the sum

$$-\frac{1}{\beta}\sum_{s_1}(\varepsilon_m - i\zeta_{s_1} - \mu)^{-1}(\varepsilon_n - i\zeta_{s_1} - \mu + i\omega_s)^{-1}, \quad \text{(C.9)}$$

which can be evaluated by the contour method discussed above. For the case $m = n$ and $s = 0$ the result is given by eq. (C.5). Otherwise, the sum (C.9) reduces to

$$(\varepsilon_n - \varepsilon_m + i\omega_s)^{-1}\left[(\mathrm{e}^{\beta(\varepsilon_m-\mu)}+1)^{-1} - (\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1)^{-1}\right]. \quad \text{(C.10)}$$

On substituting into eq. (C.8) we obtain firstly for $s \neq 0$

$$\tilde{\Gamma}^{\mathrm{E}}_{\tau\tau}(\omega_s, \boldsymbol{k}) = \frac{1}{\pi}\frac{\langle n|\mathrm{e}^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{\Pi}}}|m\rangle\langle m|\mathrm{e}^{i\boldsymbol{k}\cdot\hat{\boldsymbol{\Pi}}}|n\rangle}{\varepsilon_n - \varepsilon_m + i\omega_s}\left[\frac{1}{\mathrm{e}^{\beta(\varepsilon_m-\mu)}+1} - \frac{1}{\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1}\right],$$

and, for $s = 0$

$$\tilde{\Gamma}^{\mathrm{E}}_{\tau\tau}(0, \boldsymbol{k}) = \frac{1}{\pi}\sum_{n\neq m}\frac{\langle n|\mathrm{e}^{-ik\tilde{\Pi}}|m\rangle\langle m|\mathrm{e}^{ik\tilde{\Pi}}|n\rangle}{\varepsilon_n - \varepsilon_m}\left[\frac{1}{\mathrm{e}^{\beta(\varepsilon_m-\mu)}+1} - \frac{1}{\mathrm{e}^{\beta(\varepsilon_n-\mu)}+1}\right]$$

$$+\frac{1}{\pi}\sum_n \langle n|\mathrm{e}^{-ik\tilde{\Pi}}|n\rangle\langle n|\mathrm{e}^{ik\tilde{\Pi}}|n\rangle\frac{\beta}{4}\mathrm{sech}^2\frac{\beta}{2}(\varepsilon_n - \mu). \quad \text{(C.11)}$$

Expressions for the other components, $\tilde{\Gamma}^{\mathrm{E}}_{j\tau}$ and $\tilde{\Gamma}^{\mathrm{E}}_{jl}$ are obtained by making the

replacements

$$e^{-ik\cdot\hat{n}} \to \frac{i}{2m}\{e^{-ik\cdot\hat{n}}, \Pi_j\}, \qquad e^{ik\cdot\hat{n}} \to \frac{i}{2m}\{e^{ik\cdot\hat{n}}, \Pi_l\}, \tag{C.12}$$

as in the zero-temperature case.

The leading terms in the expansion around $k = 0$ are obtained by restricting the sum over m to the values $m = n$, $n \pm 1$, etc. and using the formula (B.15). The results are expressed through the formulae (3.16).

The low-temperature approximation is discussed in sect. 3. Here we give some of the background for computations in the high-temperature regime $\beta \ll ml^2$. When the temperature is large in comparison with the level spacing it becomes feasible to approximate the sums over discrete energy levels by integrals. For the sums appearing in eq. (3.15) it is possible to evaluate the resulting integrals in terms of elementary functions and so obtain the leading terms in expansions in powers of the small quantity β/ml^2. The coefficients in these expansions will turn out to be functions of $\beta\mu$, where μ is the chemical potential, and we need not assume that this quantity is small.

To begin with the general problem, suppose we are to approximate the sum, $\Sigma_n f_n$, where the quantities f_n are interpolated by a smooth function

$$\begin{aligned} f_n &= f(\beta\varepsilon_n) \\ &= \phi\left[\left(n + \tfrac{1}{2}\right)\alpha\right], \end{aligned} \tag{C.13}$$

$$\alpha = \frac{\beta}{ml^2}. \tag{C.14}$$

Let us further suppose that the function $\phi(z)$ is slowly varying on the scale of α

$$\alpha^2\phi''(z) \ll \phi(z),$$

etc. In order to convert the sum into an integral, it is necessary to construct a function $F(z, \alpha)$ such that

$$\phi\left[\left(n + \tfrac{1}{2}\right)\alpha\right] = \int_{n\alpha}^{(n+1)\alpha} dz F(z, \alpha). \tag{C.15}$$

With such a function we can write

$$\sum_n f_n = \int_0^\infty dz F(z, \alpha). \tag{C.16}$$

It is a simple exercise to find the first few terms in the expansion of $F(z,\alpha)$ in powers of α,

$$F(z,\alpha)=\frac{1}{\alpha}\phi(z)-\frac{\alpha}{24}\phi''(z)+\frac{7}{5760}\alpha^3\phi''''(z)+\dots . \tag{C.17}$$

One can easily verify that this function satisfies eq. (C.15) by expanding in powers of $z-(n+\frac{1}{2})\alpha$. On substituting eq. (C.17) into eq. (C.16) we obtain the formula

$$\sum_n f_n=\frac{1}{\alpha}\int_0^\infty dz\,\phi(z)+\frac{\alpha}{24}\phi'(0)-\frac{7}{5760}\alpha^3\phi'''(0)+\dots , \tag{C.18}$$

which is valid under the assumptions that the integral converges and that the interpolating function vanishes with sufficient rapidity at $z\to\infty$. These assumptions are justified in the cases to be considered.

Now, to treat the sum (3.20) which defines the chemical potential

$$\nu=\sum_n\left(e^{\beta(\varepsilon_n-\mu)}+1\right)^{-1},$$

we must choose the interpolating function

$$\phi(z)=\left(e^{z-\beta\mu}+1\right)^{-1}$$

and evaluate the quantities

$$\int_0^\infty dz\,\phi(z)=\ln(1+e^{\beta\mu}),$$

$$\phi'(0)=-\tfrac{1}{4}\operatorname{sech}^2\tfrac{1}{2}\beta\mu,$$

$$\phi'''(0)=-\tfrac{1}{4}\operatorname{sech}^2\tfrac{1}{2}\beta\mu+\tfrac{3}{8}\operatorname{sech}^4\tfrac{1}{2}\beta\mu,$$

which are to be substituted into eq. (C.18). This gives

$$\nu=\frac{1}{\alpha}\left(\beta\mu+\ln(1+e^{-\beta\mu})\right)-\tfrac{1}{96}\alpha\operatorname{sech}^2\tfrac{1}{2}\beta\mu$$
$$+\tfrac{7}{46080}\alpha^2\left(2\operatorname{sech}^2\tfrac{1}{2}\beta\mu-3\operatorname{sech}^4\tfrac{1}{2}\beta\mu\right)+\dots , \tag{C.19}$$

where $\alpha=\beta/ml^2$. If we make the further assumption that $\beta\mu$ is large, then eq.

(C.19) can be solved by iteration to give

$$\mu = \frac{\nu}{ml^2} - \frac{1}{\beta} e^{-\beta\nu/ml^2}\left(1 - \frac{1}{24}\left(\frac{\beta}{ml^2}\right)^2 + \frac{7}{5760}\left(\frac{\beta}{ml^2}\right)^4 + \ldots\right) + O(e^{-2\beta\nu/ml^2}).$$

The assumption that $\beta\mu$ is large while β/ml^2 is small requires that ν is large*. We are, in effect, obtaining an expansion in powers of $1/\nu$. On substituting for l^2 from the formula (3.19) we obtain

$$\mu = \pi\frac{n_e}{m} - \frac{1}{\beta} e^{-\beta\pi n_e/m}\left(1 - \frac{1}{24}\left(\frac{\beta\pi n_e}{\nu m}\right)^2 + \ldots\right) + \ldots. \tag{C.20}$$

Turning next to the sums which define the low-momentum behaviour of the polarization tensor, consider firstly the quantity

$$a = \frac{\beta}{4\pi l^2}\sum_n \operatorname{sech}^2 \tfrac{1}{2}\beta(\varepsilon_n - \mu) \tag{C.21}$$

Here we choose

$$\phi(z) = \operatorname{sech}^2\left(\frac{z - \beta\mu}{2}\right)$$

$$= -4\frac{d}{dz}(e^{z-\beta\mu} + 1)^{-1}.$$

To obtain the first correction term we need

$$\phi'(0) = -\sinh \tfrac{1}{2}\beta\mu \operatorname{sech}^3 \tfrac{1}{2}\beta\mu.$$

On substituting this into eq. (C.18) we obtain

$$a = \frac{m}{\pi}\left[\frac{1}{1 + e^{-\beta\mu}} - \frac{1}{96}\left(\frac{\beta}{ml^2}\right)^2 \sinh\frac{\beta\mu}{2}\operatorname{sech}^3\frac{\beta\mu}{2} + \ldots\right]. \tag{C.22}$$

Next, for the quantity

$$c = \frac{\nu}{\pi} - \frac{1}{8\pi}\frac{\beta}{ml^2}\sum_n (2n+1)\operatorname{sech}^2\frac{\beta}{2}(\varepsilon_n - \mu)$$

$$= \frac{1}{\pi}\sum_n\left[(e^{\beta(\varepsilon_n-\mu)} + 1)^{-1} - \tfrac{1}{4}\beta\varepsilon_n \operatorname{sech}^2\frac{\beta}{2}(\varepsilon_n - \mu)\right], \tag{C.23}$$

* The convergence of the expansion (C.17) necessitates that $\phi(z)$ should vary slowly on the scale of α. This implies $\beta/ml^2 \ll \beta\mu$ which is consistent with $\nu \gg 1$.

choose

$$\phi(z) = \frac{1}{\pi}\left[(e^{z-\beta\mu}+1)^{-1} - \tfrac{1}{4}z\,\mathrm{sech}^2\tfrac{1}{2}(z-\beta\mu)\right]$$

$$= \frac{1}{\pi}\frac{d}{dz}\left[z(e^{z-\beta\mu}+1)^{-1}\right],$$

$$\phi'(0) = \frac{-1}{2\pi}\mathrm{sech}^2\frac{\beta\mu}{2}, \qquad \phi'''(0) = \frac{-1}{\pi}\mathrm{sech}^2\frac{\beta\mu}{2} + \frac{3}{2\pi}\mathrm{sech}^4\frac{\beta\mu}{2}.$$

In this case the integral of $\phi(z)$ vanishes and we are left with the correction term

$$c = \frac{-1}{48\pi}\frac{\beta}{ml^2}\mathrm{sech}^2\frac{\beta\mu}{2} + \frac{7}{5760\pi}\left(\frac{\beta}{ml^2}\right)^3\left(\mathrm{sech}^2\frac{\beta\mu}{2} - \frac{3}{2}\mathrm{sech}^4\frac{\beta\mu}{2}\right) + \ldots. \quad \text{(C.24)}$$

Since, according to the definitions (3.16) and (4.7) we have $b = -ml^2c$, there remains only the quantity

$$d = \frac{1}{\pi m}\sum_n\left[(2n+1)(e^{\beta(\varepsilon_n-\mu)}+1)^{-1} - \frac{1}{16}\frac{\beta}{ml^2}(2n+1)^2\mathrm{sech}^2\frac{\beta}{2}(\varepsilon_n-\mu)\right]$$

$$= \frac{2}{\pi m}\frac{ml^2}{\beta}\sum_n\left[\beta\varepsilon_n(e^{\beta(\varepsilon_n-\mu)}+1)^{-1} - \frac{1}{8}(\beta\varepsilon_n)^2\mathrm{sech}^2\frac{\beta}{2}(\varepsilon_n-\mu)\right]. \quad \text{(C.25)}$$

Here we choose

$$\phi(z) = \frac{2}{\pi m}\frac{ml^2}{\beta}\left[z(e^{z-\beta\mu}+1)^{-1} - \frac{z^2}{8}\mathrm{sech}^2\left(\frac{z-\beta\mu}{2}\right)\right]$$

$$= \frac{1}{\pi m}\frac{ml^2}{\beta}\frac{d}{dz}\left[z^2(e^{z-\beta\mu}+1)^{-1}\right],$$

$$\phi'(0) = \frac{2}{\pi m}\frac{ml^2}{\beta}(1+e^{-\beta\mu})^{-1}, \qquad \phi'''(0) = -\frac{3}{\pi m}\frac{ml^2}{\beta}\sinh\frac{\beta\mu}{2}\mathrm{sech}^3\frac{\beta\mu}{2}.$$

Again the integral of $\phi(z)$ vanishes and we therefore obtain

$$d = \frac{1}{12\pi m}(1+e^{-\beta\mu})^{-1} + \frac{21}{5760\pi m}\left(\frac{\beta}{ml^2}\right)^2\sinh\frac{\beta\mu}{2}\mathrm{sech}^3\frac{\beta\mu}{2} + \ldots. \quad \text{(C.26)}$$

Appendix D

THE EFFECTIVE INTERACTION AT ZERO TEMPERATURE

To gain a better understanding of the dynamics of the two gauge fields, A_μ and a_μ, with their various mixing terms, it is instructive to derive the effective current–current interaction. Part of the result of this computation is exhibited in sect. 2. Here we provide some of the details.

The bilinear part of the effective action is given by

$$\Gamma = \int \mathrm{d}^3x \left[-\frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu} - \frac{\nu}{2\pi} \varepsilon^{\mu\lambda\rho} a_\mu \partial_\lambda a_\rho \right]$$

$$+ \tfrac{1}{2} \int \mathrm{d}^3x \, \mathrm{d}^3x' \, \Gamma^{\mu\nu}(x-x') \big(A_\mu(x) + a_\mu(x)\big)\big(A_\nu(x') + a_\nu(x')\big), \quad \text{(D.1)}$$

where the Fourier components of $\Gamma_{\mu\nu}$ are expressed in terms of the three invariants, Π_0, Π_1 and Π_2, which are constants in the long-wavelength, low-frequency approximation (see eqs. (2.14) and (2.15)).

To obtain the effective interaction between a pair of conserved electromagnetic currents we must solve the field equations,

$$\delta\Gamma/\delta A_\mu(x) = -I^\mu(x), \qquad \delta\Gamma/\delta a_\mu(x) = 0, \quad \text{(D.2)}$$

and evaluate the amplitude

$$W(I) = \Gamma(A,a) + \int \mathrm{d}^3x I^\mu(x) A_\mu(x)$$

$$= \tfrac{1}{2} \int \mathrm{d}^3x I^\mu(x) A_\mu(x)$$

$$= \tfrac{1}{2} \int \mathrm{d}^3x \, \mathrm{d}^3x' \, I^\mu(x) D_{\mu\nu}(x-x') I^\nu(x'). \quad \text{(D.3)}$$

In Fourier space the field equations (D.2) are

$$-\frac{1}{e^2} k_\nu (k^\nu \tilde{A}^\mu - k^\mu \tilde{A}^\nu) + \tilde{\Gamma}^{\mu\nu}(k)\big(\tilde{A}_\nu + \tilde{a}_\nu\big) = -\tilde{I}^\mu,$$

$$-\frac{i\nu}{\pi} \varepsilon^{\mu\lambda\rho} k_\lambda \tilde{a}_\rho + \tilde{\Gamma}^{\mu\nu}(k)\big(\tilde{A}_\nu + \tilde{a}_\nu\big) = 0, \quad \text{(D.4)}$$

where $k^\mu \tilde{I}_\mu = \omega \tilde{I}_0 + k_j \tilde{I}_j = 0$. Since we do not generally have Lorentz invariance

(which is obtained only for $\Pi_2 = -\Pi_0$) we cannot choose any particularly advantageous frame. However, it is helpful to employ a complex basis, writing

$$\tilde{A}_\pm = \frac{1}{\sqrt{2}}(\tilde{A}_1 \pm i\tilde{A}_2),$$

etc. Since eq. (D.4) is gauge invariant we can look for the solution with

$$\tilde{A}_0 = \tilde{a}_0 = 0. \tag{D.5}$$

With this restriction the equations reduce to the form

$$\frac{1}{e^2}\left[(\omega^2 - k_+k_-)\tilde{A}_\pm + k_\pm^2\tilde{A}_\mp\right] + \Gamma_{\pm\mp}(\tilde{A}_\pm + \tilde{a}_\pm) + \Gamma_{\pm\pm}(\tilde{A}_\mp + \tilde{a}_\mp) = -I_\pm$$

$$\mp\frac{\nu}{\pi}\omega a_\pm + \Gamma_{\pm\mp}(\tilde{A}_\pm + a_\pm) + \Gamma_{\pm\pm}(\tilde{A}_\mp + \tilde{a}_\mp) = 0. \tag{D.6}$$

On eliminating $a_\pm$ they reduce to

$$\alpha_+\tilde{A}_+ + \beta_+\tilde{A}_- = -j_+,$$

$$\beta_-\tilde{A}_+ + \alpha_-\tilde{A}_- = -j_-, \tag{D.7}$$

where

$$j_+ = \left(1 - \frac{\pi}{\nu}\Pi_1 - \frac{\pi}{\nu}\omega\Pi_0\right)\tilde{I}_+ + \frac{\pi}{\nu}\Pi_2 k_+\tilde{I}_0,$$

$$j_- = \left(1 - \frac{\pi}{\nu}\Pi_1 + \frac{\pi}{\nu}\omega\Pi_0\right)\tilde{I}_- - \frac{\pi}{\nu}\Pi_2 k_-\tilde{I}_0, \tag{D.8}$$

and the coefficients $\alpha_\pm, \beta_\pm$ are given by

$$\alpha_\pm = \mp\frac{\pi}{\nu e^2}\omega\left(\omega^2\Pi_0 - \frac{\nu}{\pi}e^2\Pi_1 + k_+k_-(\Pi_2 - \Pi_0)\right)$$

$$+ \frac{\omega^2}{e^2}\left(1 - \frac{\pi}{\nu}\Pi_1 + e^2\Pi_0\right) - \frac{k_+k_-}{e^2}\left(1 - \frac{\pi}{\nu}\Pi_1 - e^2\Pi_2\right),$$

$$\beta_\pm = \mp\frac{\pi}{\nu e^2}k_\pm^2\omega(\Pi_0 + \Pi_2) + \frac{k_\pm^2}{e^2}\left(1 - \frac{\pi}{\nu}\Pi_1 - e^2\Pi_2\right). \tag{D.9}$$

The solution is

$$\tilde{A}_+ = \frac{1}{\omega^2 \Delta}(\alpha_- j_+ - \beta_+ j_-),$$

$$\tilde{A}_- = \frac{1}{\omega^2 \Delta}(\alpha_+ j_- - \beta_- j_+), \tag{D.10}$$

where the denominator is given by

$$\begin{aligned}\Delta &= -\frac{1}{\omega^2}(\alpha_+\alpha_- - \beta_+\beta_-) \\ &= \left(\frac{\pi}{\nu e^2}\right)^2\left[\left\{(\omega^2 - 2k_+k_-)\Pi_0 - \frac{\nu}{\pi}e^2\Pi_1\right\}^2 - \left(\frac{\nu}{\pi}\right)^2(\omega^2 - 2k_+k_-)\left(1 - \frac{\pi}{\nu}\Pi_1 + e^2\Pi_0\right)^2\right. \\ &\quad \left. + 2k_+k_-(\Pi_0 + \Pi_2)\left\{(\omega^2 - 2k_+k_-)\Pi_0 - \left(\frac{\nu}{\pi}\right)^2 e^2(1 + e^2\Pi_0)\right\}\right]. \end{aligned} \tag{D.11}$$

This expression is quadratic in ω^2 which implies that there are two branches to the dispersion relation. Their values at $k = 0$ define the "masses" $\mu_\pm^2$, referred to in the text. They are real and positive if $e^2\Pi_0$ is positive

$$\mu_\pm^2 = \frac{1}{4}\left(\frac{\nu}{\pi}\right)^2 \frac{1}{\Pi_0^2}\left[\left(1 - \frac{\pi}{\nu}\Pi_1 + e^2\Pi_0\right) \pm \sqrt{\left(1 - \frac{\pi}{\nu}\Pi_1 - e^2\Pi_0\right)^2 + 4e^2\Pi_0}\right]^2. \tag{D.12}$$

The solutions (D.10) are singular at $\omega = 0$ and one might suspect that this is merely a gauge artifact. However, this may not be the case, at least in higher orders. A careful evaluation of the functional (D.3) shows that the singularity may persist and give rise to a long-range instantaneous interaction between the currents. The final result is

$$\begin{aligned} W(I) &= \tfrac{1}{2}\int \frac{d\omega\, d^2k}{(2\pi)^3}\frac{1}{e^2\Delta}\left[\left\{\left(1 - \frac{\pi}{\nu}\Pi_1\right)^2 + e^2\Pi_0 - \left(\frac{\pi}{\nu}\right)^2\Pi_0^2(\omega^2 - k^2)\right.\right. \\ &\quad \left. - \frac{1}{2}\left(\frac{\pi}{\nu}\right)^2\Pi_0(\Pi_0 + \Pi_2)k^2\right\} I_i^* I_i - \left\{\left(1 - \frac{\pi}{\nu}\Pi_1\right)^2 - e^2\Pi_2\right. \\ &\quad \left. + \left(\frac{\pi}{\nu}\right)^2\Pi_0\Pi_2(\omega^2 - k^2)\right\} I_0^* I_0 - \left\{\frac{\pi}{\nu}\Pi_1\left(1 - \frac{\pi}{\nu}\Pi_1\right) + \left(\frac{\pi}{\nu}\right)^2\Pi_0^2(\omega^2 - k^2)\right. \\ &\quad \left.\left. + \left(\frac{\pi}{\nu}\right)^2\Pi_0(\Pi_0 + \Pi_2)k^2\right\}\frac{\nu e^2}{\pi}\frac{i}{\omega}\varepsilon_{ij} I_i^* I_j\right]. \end{aligned} \tag{D.13}$$

The combination of currents that appears in the last term is in fact Lorentz invariant. On using current conservation it can be expressed in the Chern–Simons form

$$-\frac{1}{\omega}\varepsilon_{ij}I_i^*I_j = \frac{1}{k^2-\omega^2}\varepsilon^{\mu\lambda\rho}I_\mu^* k_\lambda I_\rho$$

$$= \frac{1}{k^2}\varepsilon_{ij}\left(I_0^* k_i I_j - I_j^* k_i I_0\right). \tag{D.14}$$

Indeed, the entire functional (D.13) becomes Lorentz invariant, as it should, when $\Pi_2 = -\Pi_0$. If we substitute the 1-loop value for Π_1, as given in eq. (2.15), the residue of the pole at $\omega = 0$ vanishes. However, this may not be the case in higher orders.

References

[1] W. Siegal, Nucl. Phys. B156 (1979) 135;
J. Schonfeld, Nucl. Phys. B185 (1981) 157;
S. Deser, R. Jackiw and S. Templeton, Phys. Rev. Lett. 48 (1983) 975; Ann. Phys. 140 (1982) 372

[2] Y. Hosotani, Phys. Rev. Lett. 62 (1989) 2785;
S. Elitzur, G. Moore, A. Schwimmer and N. Seiberg, IAS preprint, IASSNS-HEP-89/20 (1989)

[3] V. Kalmeyer and R.B. Laughlin, Phys. Rev. Lett. 59 (1987) 2095;
R.B. Laughlin, Phys. Rev. Lett. 60 (1988) 1057;
I. Dzyaloshinskii, A.M. Polyakov and P. Wiegmann, Phys. Lett. A127 (1988) 112

[4] Yi-Hong Chen, F. Wilczek, E. Witten and B.I. Halperin, Int. J. Mod. Phys. B3 (1989) 1001

[5] F. Wilczek and A. Zee, Phys. Rev. Lett. 51 (1983) 2250;
Yong-Shi Wu and A. Zee, Phys. Lett. B147 (1984) 325;
D.P. Arovas, R. Schrieffer, F. Wilczek and A. Zee, Nucl. Phys. B251 (1985) 117

[6] F. Wilczek, Phys. Rev. Lett. 49 (1982) 957

[7] J.M. Leinaas and J. Myrheim, Nuovo Cim. B37 (1977) 1;
F. Wilczek, Phys. Rev. Lett. 48 (1982) 1144

[8] Y. Hosotani and S. Chakravarty, IAS preprint, IASSNS-HEP-89/31 (1989)

[9] E.M. Lifshitz and L.P. Pitaevskii, Statistical Physics, part 2 (Pergamon Press, Oxford, 1980)

Volume 240, number 1,2 PHYSICS LETTERS B 19 April 1990

ANYONS AND CHERN–SIMONS THEORY ON COMPACT SPACES OF FINITE GENUS

S. RANDJBAR-DAEMI [a], Abdus SALAM [a,b] and J.A. STRATHDEE [a]

[a] *International Centre for Theoretical Physics, I-34100 Trieste, Italy*
[b] *Imperial College, Prince Consort Road, London SW7 2BZ, UK*

Received 30 January 1990

We study the coupling of an abelian Chern–Simons field to fermions in space–times of the form $\mathbb{R}\times M^2$, where M^2 is a compact riemannian manifold. Upon integrating out the non-zero modes of the Chern–Simons field, an effective N-particle hamiltonian is constructed, which involves a term representing the effects of the zero modes. We also study the transformation to the fractional statistics (anyon) basis. It is shown that unlike the case of the flat euclidean M^2 the anyon wave equation involves some residual metric dependent interactions, and the wave function is multivalued.

1. Fractional statistics and Chern–Simons theory

The (2+1)-dimensional Chern–Simons theories have been the subject of many recent studies [1]. It is well known that the abelian gauge field in (2+1) dimensions is without any dynamical degrees of freedom [2]. It can be solved for and eliminated, leaving only a characteristic instantaneous interaction between the charged particles to which it couples [3]. This is true at least when the spacetime is topological trivial. However, if there are nontrivial homotopies then it is possible for the Chern–Simons field to carry zero mode excitations which correspond to true dynamical degrees of freedom: they give rise, in effect, to retarded interactions among the charged particles [4]. In this note we shall consider spacetimes of the form $M^2\times\mathbb{R}$ where M^2 is a compact riemannian space and $\mathbb{R}$ is the time. Since the Chern–Simons action functional does not depend on the metric, we can regard M^2 as a Riemann surface and employ the machinery of conformal geometry to analyze the theory. In such cases we can integrate out the Chern–Simons field, apart from the zero mode variables, and express the resulting instantaneous interaction in terms of a scalar Green's function, which is given in terms of well known Riemann surface quantities such as the prime form. The zero mode couplings, on the other hand, are expressed in terms of the holomorphic one-forms of the Riemann surface.

One of the motivations for studying abelian Chern–Simons theory in (2+1) dimensions is the intimate association with fractional statistics [5]. In particular, it is possible to construct a (singular) gauge transformation which converts a system of ordinary fermions with Chern–Simons interactions into a system of free particles which obeys fractional statistics (anyons) [6]. This is true only if M^2 has the topology of $\mathbb{R}^2$. Otherwise the anyons are not free. There is generally a residual, metric dependent background and, in the case of higher genus, there is a coupling to the zero modes and a residual interaction between the anyons which cannot be gauged away. The purpose of this note is to give a formal description of these residual effects in the case of a simple model in which the Chern–Simons field is coupled to non-relativistic fermions on a compact surface without boundaries.

Reprinted with permission from *Phys. Lett.* **240** (1990) 121–126
© 1990 Elsevier Science Publishers B.V.

2. The model

The action functional describing a system of non-relativistic charged fermions, ψ, to a Chern–Simons field, a_μ, is given by

$$S=\int \mathrm{d}^3x\left(\sqrt{g}\,\psi^+\,\mathrm{i}\nabla_0\psi-\frac{1}{2m}\sqrt{g}\,g^{ij}\nabla_i\psi^+\nabla_j\psi-\frac{k}{4\pi}\varepsilon^{\lambda\mu\rho}a_\lambda\partial_\mu a_\rho\right), \tag{1}$$

where g_{ij} is a time independent metric on M^2 and $\varepsilon^{\lambda\mu\rho}$ is the three-dimensional permutation symbol. We shall not treat the metric as a dynamical variable. The fermions may carry spin in which case the covariant derivative, ∇_i would have to include the spin connection. For simplicity we shall here treat the fermions as scalars. The covariant derivatives therefore contain only the Chern–Simons connection

$$\nabla_\mu\psi=(\partial_\mu-\mathrm{i}a_\mu)\psi,\quad \nabla_\mu\psi^+=(\partial_\mu+\mathrm{i}a_\mu)\psi^+, \tag{2}$$

The action functional (1) is clearly invariant with respect to infinitesimal, or topologically trivial gauge transformations. However, as is well known, it is not invariant with respect to the topologically non-trivial transformations: it is shifted by integer multiples of $2\pi k$. The coupling parameter, k, is required to be an integer in order that $\exp(\mathrm{i}S)$, and hence the quantized theory, shall be invariant.

The equations of motion for the Chern–Simons field take the form

$$\varepsilon^{\lambda\mu\rho}\partial_\mu a_\rho=-\frac{2\pi}{k}J^\lambda, \tag{3}$$

which implies that the field strength must vanish in empty space. It follows that the Chern–Simons field can be gauged away from regions where no charged particles are located. The transformation which achieves this is necessarily singular at the particle locations and it leads to the multivalued wave functions characteristic of fractional statistics [5,3]. One of our aims is to specify this transformation in the case of compact manifolds of finite genus. The strategy is to integrate out the Chern–Simons field to obtain the effective action for the charged fields ψ and ψ^+. From this we can obtain the N-fermion hamiltonian with coupling effects included. The gauge transformation which eliminates these couplings – up to residual effects – can then be found by inspection.

3. The effective action

Elimination of the Chern–Simons field is a canonical exercise which can be carried out most conveniently in the temporal gauge, $a_0=0$. However, owing to the compact and homotopically non-trivial features of the space, it is necessary to exercise some care, and we therefore sketch the procedure [#1].

Since the abelian Chern–Simons theory is gaussian we have

$$\int(\mathrm{d}a)\delta(a_0)\exp\left[\mathrm{i}\int\mathrm{d}^3x\left(-\frac{k}{4\pi}\varepsilon^{\lambda\mu\rho}a_\lambda\partial_\mu a_\rho-J^\mu a_\mu\right)\right]=\exp\left(-\frac{\mathrm{i}}{2}\int\mathrm{d}^3x\,J^\mu a_\mu^{\mathrm{cl}}\right), \tag{4}$$

where a_μ^{cl} is a particular solution of the classical equations (3). The result (4) should be independent of the gauge choice if the current, J^μ, is conserved. This will be the case if the effective action is used for the computation of on-shell quantities. In temporal gauge the appropriate classical solution is quite easily expressed,

$$a_0^{\mathrm{cl}}=0,\quad a_z^{\mathrm{cl}}=-\frac{2\pi\mathrm{i}}{k}\int^t\mathrm{d}t'\,J_z'. \tag{5}$$

[#1] Details will be presented elsewhere.

Here, and in the following we use conformal coordinates on M^2. The line element is given by $ds^2=2\rho\, dz\, d\bar{z}$ where $\rho(z)$ is real and positive. Although the effective interaction (4), bilinear in the current when a^{cl} is given by (5), is independent of the metric, we find it is necessary to use a metric-containing Green's function in order to separate the instantaneous (non-dynamical) effects from the zero-mode (retarded) contributions. To describe this Green's function we must recall some notions from conformal geometry [7].

On a Riemann surface of finite genus, γ, we can define the multivalued abelian integrals

$$\varphi_i(z)=\int^z dw\,\omega_i(w),\quad i=1,\ldots,\gamma, \tag{6}$$

where the ω_i are holomorphic one-forms. We also have the prime form, $E(z, w)$, which is holomorphic and multi-valued. It is antisymmetric and, near $z=w$, $E(z, w)=z-w+\ldots$. Out of these we can make the quantity

$$F(z, w)=|E(z, w)|^2\exp\{-2\pi[\varphi_2(z)-\varphi_2(w)]\tau_2^{-1}[\varphi_2(z)-\varphi_2(w)]\}$$

which is real, symmetric and single-valued. Here $\varphi_{2i}(z)$ denotes the imaginary part of $\varphi_i(z)$ and $(\tau_2)_{ij}$ denotes the imaginary part of the period matrix, τ_{ij}. The quantities φ, E and F are conformal in the sense that they depend on the conformal class of the metric but not on the scale factor, $\rho(z)$. This is in contrast with the Green's function defined by [7]

$$G(z, w)=-\ln F(z, w)+\int d^2x\frac{\rho(x)}{A}[\ln F(z, x)+\ln F(x, w)]-\int d^2x\, d^2y\frac{\rho(x)\rho(y)}{A^2}\ln F(x, y), \tag{7}$$

which depends explicitly on ρ (and the area $A=\int d^2z\,\rho$) and is therefore not conformal. However, the Green's function is real, symmetric and single-valued, and it transforms as a scalar with respect to analytic reparametrizations. Most important for our purposes, it satisfies the differential equations

$$\partial_z\partial_{\bar{z}}G(z, w)=-2\pi\delta_2(z, w)+2\pi\frac{\rho(z)}{A},\quad \partial_z\partial_w G(z, w)=2\pi\delta_2(z, w)-\pi\omega(z)\tau_2^{-1}\overline{\omega(w)}. \tag{8a,b}$$

Using these, the effective interaction can be expressed as follows:

$$\begin{aligned}-\tfrac{1}{2}\int d^2z\, dt(J_z a_z^{cl}+J_{\bar{z}}a_{\bar{z}}^{cl})&=\frac{i\pi}{k}\int d^2z\, dt\, dt'[J_z\theta(t-t')J_z'-J_{\bar{z}}\theta(t-t')J_{\bar{z}}']\\&=\frac{i\pi}{k}\int d^2z\, d^2w\, dt\, dt'\, J_{\bar{z}}\varepsilon(t-t')\delta_2(z, w)J_w'\\&=\frac{i}{2k}\int d^2z\, d^2w\, dt\, dt'[\partial_z J_{\bar{z}}\varepsilon(t-t')G(z, w)\partial_w J_w'+\pi J_{\bar{z}}\omega(z)\tau_2^{-1}\varepsilon(t-t')\overline{\omega(w)}J_w'],\end{aligned} \tag{9}$$

where $\theta(t-t')$ is the step function, etc. We have used (8b) to introduce $G(z, w)$ into the integrand. With the imposition of current conservation together with suitable assumptions, about the asymptotic current configurations, it is possible to express the right-hand side of (9) in the form

$$\frac{i}{2k}\int d^2z\, d^2w\, dt\, J_0(z, t)G(z, w)(\partial_w J_w-\partial_w J_w)+\frac{i\pi}{2k}\int dt\, dt'\,\bar{Q}(t)\tau_2^{-1}\varepsilon(t-t')Q(t'), \tag{10}$$

where $Q_i(t)$ denotes the source component that couples to the zero modes.

$$Q_i(t)=\int d^2z\,\overline{\omega_i(z)}J_z. \tag{11}$$

By introducing a set of dynamical variables $u_i(t)$, $\bar{u}_i(t)$, $i=1,\ldots,\gamma$ with the 2-point function,

$$\langle \mathrm{T}\, u_i(t)\bar{u}_j(t')\rangle = \frac{\pi}{k}(\tau_2^{-1})_{ij}\varepsilon(t-t')\,, \tag{12}$$

we can simulate the zero mode exchange,

$$\exp\left(-\frac{\pi}{2k}\int \mathrm{d}t\,\mathrm{d}t'\,\bar{Q}(t)\tau_2^{-1}\varepsilon(t-t')Q(t')\right) = \left\langle \exp\left(\frac{1}{\mathrm{i}}\int \mathrm{d}t(\bar{u}Q+\bar{Q}u)\right)\right\rangle . \tag{13}$$

The other term in (10) describes an instantaneous interaction between the currents. Taken together these terms define the Chern–Simons induced interactions between the fermions. It can be shown [#2] that they are governed by the effective action functional

$$S=\int \mathrm{d}t\left[\frac{k}{\pi}\bar{u}\tau_2\mathrm{i}\partial_0 u+\int \mathrm{d}^2z\left(\rho\psi^+\mathrm{i}\partial_0\psi-\frac{1}{2m}(\nabla_z\psi^+\nabla_z\psi+\nabla_z\psi^+\nabla_z\psi)\right)\right], \tag{14}$$

where the covariant derivatives are of the form (2) but with the connection

$$a_z=\omega(z)\cdot u+\frac{\mathrm{i}}{2k}\int \mathrm{d}^2w\,\partial_z G(z,w)J_0(w)\,, \tag{15}$$

where the charge density is given by

$$J_0=\rho\psi^+\psi\,. \tag{16}$$

4. The N-particle hamiltonian

From the action integral (14) it is straightforward to extract the N-particle hamiltonian. Define the N-fermion basis vectors $\langle u|\psi(\xi^1)...\psi(\xi^N)$ where $\langle u|\psi^+(\xi)=0$. The vacua are chosen to diagonalize the "holomorphic" zero mode operators u_i, the conjugate operators are given by

$$\langle u|\bar{u}_i=-\frac{\mathrm{i}\pi}{k}(\tau_2^{-1})_{ij}\frac{\partial}{\partial u_j}\langle u|\,. \tag{17}$$

The action of the hamiltonian on the N-fermion states is given by

$$\langle u|\psi(\xi^1)...\psi(\xi^N)H=\langle u|\psi(\xi^1)...\psi(\xi^N)\int \mathrm{d}^2z\frac{1}{2m}(\nabla_z\psi^+\nabla_z\psi+\nabla_z\psi^+\nabla_z\psi)$$

$$=\sum_{\alpha=1}^{N}\left(-\frac{\rho(\xi^\alpha)^{-1}}{2m}\right)(\nabla_\alpha\nabla_{\bar{\alpha}}+\nabla_{\bar{\alpha}}\nabla_\alpha)\langle u|\psi(\xi^1)...\psi(\xi^N)\,, \tag{18}$$

where the covariant derivative in N complex dimensions is defined by

$$\nabla_\alpha=\frac{\partial}{\partial\xi^\alpha}-\mathrm{i}A_\alpha(\xi),\quad \nabla_{\bar{\alpha}}=\frac{\partial}{\partial\bar{\xi}^\alpha}-\mathrm{i}A_{\bar{\alpha}}(\xi),\quad A_\alpha(\xi)=\omega(\xi^\alpha)\cdot u+\frac{\mathrm{i}}{2k}\sum_{\beta\neq\alpha}\partial_\alpha G(\xi^\alpha,\xi^\beta)\,, \tag{19}$$

and likewise for the components $A_{\bar{\alpha}}=(A_\alpha)^*$. The hamiltonian is thereby reduced to a kind of N-dimensional Laplace operator. It is covariant with respect to the analytic reparametrizations $\zeta^\alpha\to f(\xi^\alpha)$, $\alpha=1,...,N$ and also with respect to N-dimensional gauge transformations, $A_\alpha\to A_\alpha+\partial_\alpha\Lambda$. (It should be pointed out that the result

[#2] There is a technical complication arising from the $a^2\psi^+\psi$ interaction in (1). To avoid this it is necessary to introduce auxiliary fermionic variables $\Pi\sim\nabla\psi$ so that the currents do not depend explicitly on a. The auxiliary variables can be eliminated after a itself has been eliminated.

(18), (19) is exact. There is no need to regularize the short-distance behaviour of $G(z, w)$ because the sum (19) automatically excludes terms with $\beta=\alpha$.) It remains to consider how much of the interaction can be gauged away.

5. Fractional statistics

The transition to a new (anyonic) basis in which the N-particle wave functions exhibit fractional statistics is effected by means of a singular gauge transformation. In the original basis the interactions are described by the complex N-vector, A_α, which has singularities on the complex lines, $\xi^1=\xi^2$, etc., and we can expect the gauge transformation to have these singularities. Away from these singular lines the components of the field strength corresponding to (19) are given by

$$F_{\alpha\beta}=0, \quad F_{\alpha\beta}=\frac{i\pi}{k}\omega(\xi^\alpha)\tau_2^{-1}\overline{\omega(\xi^\beta)} \quad \beta\neq\alpha, \quad F_{\alpha\beta}=-\frac{i2\pi}{k}(N-1)\frac{\rho(\xi^\alpha)}{A} \quad \beta=\alpha, \tag{20}$$

where we have used (8). This structure is to be contrasted with the planar case (genus $\gamma=0$, area $A=\infty$) where all components vanish. Here we shall not be able to gauge away the interaction so as to arrive at a system of free anyons. The best that can be achieved is to write

$$A_\alpha=\mathscr{A}_\alpha+\partial_\alpha\Omega, \tag{21}$$

where Ω is real, defined so as to remove the singular parts of the connection (19). We choose

$$\Omega=\sum_\alpha\frac{\pi}{k}(N-1)\eta(\xi^\alpha)\tau_1\eta(\xi^\alpha)+\sum_{\alpha<\beta}\left[\frac{1}{2ik}\ln\left(\frac{E(\xi^\alpha,\xi^\beta)}{E(\xi^\alpha,\xi^\beta)^*}\right)-\frac{2\pi}{k}\eta(\xi^\alpha)\tau_1\eta(\xi^\alpha)\right], \tag{22}$$

where $\tau_1=\operatorname{Re}\tau$ and $\eta=\tau_2^{-1}\operatorname{Im}\varphi$. The corresponding residual connection takes the form

$$\mathscr{A}_\alpha=b(\xi^\alpha)-\frac{i\pi}{k}\sum_{\beta\neq\alpha}\omega(\xi^\alpha)\tau_2^{-1}\bar{\tau}\eta(\xi^\beta), \tag{23}$$

in which b is a metric-dependent background given by

$$b(z)=\frac{i\pi}{2k}(N-1)\omega(z)\tau_2^{-1}\eta(z)+\omega(z)\cdot u+\frac{i}{2k}(N-1)\int d^2w\frac{\rho(w)}{A}\left[\partial_z\ln E(z,w)-2\pi i\omega(z)\eta(w)\right]. \tag{24}$$

The expressions (22)–(24) are somewhat complicated. In fact $\mathscr{A}_\alpha$ is a k-valued vector potential on M^2. For example, on taking the coordinate ξ^β around the cycle $m\cdot a+n\cdot\beta$ the change in $\mathscr{A}_\alpha$ is given by a gauge transformation,

$$\delta_{mn}^{(\beta)}\mathscr{A}_\alpha=\partial_\alpha\Lambda_{mn}^{(\beta)}, \quad \Lambda_{mn}^{(\beta)}=\frac{2\pi}{k}(N-1)\mu(\xi^\beta)\cdot n+\frac{2\pi}{k}\sum_{\alpha\neq\beta}\mu(\xi^\alpha)\cdot n. \tag{25}$$

In these formulae real functions, $\mu(z)$ are defined like $\eta(z)$, in terms of the abelian integrals (6),

$$\varphi=\mu+\tau\eta, \quad \bar{\varphi}=\mu+\bar{\tau}\eta. \tag{26}$$

They have simple periods,

$$\delta_{mn}\mu_i=m_i, \quad \delta_{mn}\nu_i=n_i. \tag{27}$$

It follows that $\delta_{mn}^{(\alpha)}\Lambda_{mn}^{(\beta)}=(2\pi/k)(N\delta^{\alpha\beta}-1)n\cdot m'$ and hence $\exp(ik\Lambda_{mn}^{(\beta)})\in U(1)$. Thus the wave functions are k-valued.

The wavefunctions in the new basis are obtained from the original ones by multiplication with the phase factor,

$$\exp(-i\Omega)=\prod_{\alpha<\beta}\left(\frac{E(\xi^\alpha,\xi^\beta)}{E(\xi^\alpha,\xi^\beta)^*}\right)^{-1/2k}\exp\left(-\frac{i\pi}{k}(N-1)\sum_\alpha\eta(\xi^\alpha)\tau_1\eta(\xi^\alpha)+\frac{2\pi i}{k}\sum_{\alpha<\beta}\eta(\xi^\alpha)\tau_1\eta(\xi^\beta)\right), \tag{28}$$

which is clearly multivalued with respect to particle interchange. For example, with $\xi^1-\xi^2\to\exp(i\pi)(\xi^1-\xi^2)$ we obtain

$$\exp(-i\Omega)\to\exp(-i\Omega)\exp(-i\pi/k)\,, \tag{29}$$

on using the antisymmetry of $E(\xi^1,\xi^2)$. This is the usual anyon phase factor [5,6] now generalized to the finite genus cases. The new feature here is that the anyon hamiltonian is not free, but includes interactions in the form of a residual connection, $\mathscr{A}_\alpha$ specified by (23) and (24).

References

[1] E. Witten, Commun. Math. Phys. 121 (1989) 351;
M. Bos and V.P. Nair, Columbia preprints CU-TP 425, CU-TP 432;
G.V. Dunne and C.A. Trugenberger, Mod. Phys. Lett. A 4 (1989) 1635;
G.V. Dunne, R. Jackiw and C.A. Trugenberger, MIT preprint CTP 1711 (1989);
T. Killingbeck, Cambridge preprint DAMPT/88-31;
S. Elitzer, G. Moore, A. Schwimmer and N. Seiberg, IAS preprint IASSNS-HEP-89/20;
A.P. Polychronakos, University of Florida preprints UFIFT-HEP-89-7, UFIFT-HEP-89-9;
A.P. Balachandran, M. Bourdeau and S. Jo, Syracuse preprint SU-4228-402.
[2] C. Hagen, Ann. Phys. (NY) 157 (1984) 342.
[3] Y.-Hong Chen, F. Wilczek, E. Witten and B.I. Halperin, IAS preprint IASSNS-HEP-89/27 (1989).
[4] Y. Hosotani, Phys. Rev. Lett. 62 (1989) 2785.
[5] F. Wilczek and A. Zee, Phys. Lett. 51 (1983) 2250;
Y.-S. Wu and A. Zee, Phys. Lett. B 147 (1984) 325;
D.P. Arovas, R. Schrieffer, F. Wilczek and A. Zee, Nucl. Phys. B 251 (1985) 117.
[6] J.N. Leinaas and J. Myrheim, Nuovo Cimento 37B (1977) 1;
F. Wilczek, Phys. Rev. Lett. 48 (1982) 1144; 49 (1982) 957;
for a recent review see R. Mackenzie and F. Wilczek, Intern. J. Mod. Phys. A 3 (1988) 2827.
[7] L. Alvarez-Gaumé, G. Moore and C. Vafa, Commun. Math. Phys. 106 (1986) 1;
M.A. Namazie, K.S. Narain and M.H. Sarmadi, Phys. Lett. B 177 (1986) 329;
E. Verlinde and H. Verlinde, Nucl. Phys. B 288 (1987) 357.

J Mol Evol (1991) 33:105–113

Journal of Molecular Evolution

© Springer-Verlag New York Inc. 1991

The Role of Chirality in the Origin of Life

Abdus Salam

International Centre for Theoretical Physics, Trieste, Italy, and Department of Theoretical Physics, Imperial College, London, United Kingdom

"Any man who, upon looking down at his bare feet, doesn't laugh, has either no sense of symmetry or no sense of humour" (Descartes, cf. Walker 1979)

Summary. The role of chirality in the theories that determine the origin of life are reemphasized—in particular the fact that almost all amino acids utilized in living systems are of the L type. Starting from Z^0 interactions, I speculate on an explanation of the above fact in terms of quantum mechanical cooperative and condensation phenomena (possibly in terms of an *e–n* condensate where the *e–n* system has the same status as Cooper-pairing), which could give rise to second-order phase transitions (including D to L transformations) below a critical temperature T_c. As a general rule, T_c is a low temperature. From this, it is conceivable that the earth provided too hot a location for the production of L amino acids. I suggest laboratory testing of these ideas by looking for the appropriate phase transitions.

Key words: Prebiotic chirality — Origin of life — Condensation

Section 1

One may summarize the presently accepted view of the origin of life as occurring in three stages: the cosmic stage; the prebiotic chemical stage, and the biological stage:

1) The cosmic stage concerns itself with the early history of the universe where electroweak forces made a phase transition into two forces, electromagnetic and weak, 10^{-12} s after the universe was born. The temperature was then 250 GeV and the carriers of the neutral weak force—the Z_0 particle—acquired mass.

2) Chemistry became important after the planets were formed (some 10 billion years later), though it may have played a role in the presolar epochs as well (long after the quarks of the early cosmic era had condensed into protons and neutrons and much after the recombination with electrons, which took place some 10^5 years following the big bang). Molecules of future life could thus have formed even before the origin of the Earth itself (Orò et al. 1990a).

3) The biological era concerns itself with the replication of nucleic acid polymers and protein synthesis. The biological era may have started some 3.8 billion years ago.

Section 2

Classically, a chiral molecule and its mirror image [defined by left (L) or right (D) optical/rotatory dispersion] have been considered energetically equivalent. However, the parity-violating weak interactions give rise to L and D configurations (Mason and Tranter 1984), and ensure that this equivalence is no longer exact—one of the two molecules, L or D, being energetically stabilized, with energy differences on the order of 3×10^{-19} eV.

In living systems, protein molecules are composed of 20 L amino acids (although some amino acids of the opposite D type do occur in cell walls of certain bacteria). Of the 74 amino acids, for example, found in samples of the Murchison meteorite, only 8 are present in proteins, 11 have other biological roles, and the remaining 55 have been found only in extraterrestrial samples (Knervolden et al. 1971; Cronin 1989).

106

The polynucleotides contain sugars in D configurations only. Clearly, once living processes had selected handedness, the complex machinery of protein synthesis and stereoselectivity of enzymes could have assured that such handedness was perpetuated.

Section 3

The most significant of the parity-violating weak interactions are the weak–neutral, these being mediated by the Z^0 bosons (Harris et al. 1978). These interactions are of exceptionally short range by atomic and strong nuclear physics standards, and to a very good approximation may be taken as contact phenomena. Recent calculations (Mason and Tranter 1984; Tranter and MacDermott 1989) indicate that four of the amino acids in aqueous zwitterionic conformation—essentially all the ones for which these calculations are available: alanine, valine, serine, and aspartic acid—are L stabilized relative to their unnatural D mirrors for configurations in aqueous media. (A minor triumph, for the sugars—particularly for D-glyceraldehyde—the calculations show that the right-handed variety is the more stable.)

This stability affects 1 out of 10^{17} molecules at room temperatures (since $10^{-17} \approx (3 \times 10^{-19}\ \text{eV})/300°\text{K})k_B$. It is the smallness of this figure that has prompted many chemists to wonder if this mechanism could indeed be responsible for the ultimate optical asymmetry.

The crucial problem is that of amplification of this electroweak advantage over the course of time so that, for example, the 20 amino acids (which make up the proteins) convert almost entirely from D into L types. This problem has been considered by Kondepudi and Nelson (1985) following the seminal ideas of F.C. Franck.

Consider, for example, quartz crystals that may be taken as nonequilibrium statistical mechanical systems at ambient temperatures on the order of $T = 300°\text{K}$. Using an autocatalytic mechanism and theory of delayed bifurcations, one can show that a lake 1 km^2 and 4 m deep would need 10^4 years to produce the necessary electroweak advantage so far as quartz is concerned. Kondepudi and Nelson give a general theory of spontaneous chiral symmetry breaking in nonequilibrium chemical systems and the possible influence of weak–neutral currents in such a process. They conclude that on a long time scale (10^4 years), for reactions occurring in large volumes, such as the oceans, the effects of parity violations due to weak–neutral currents cannot be considered small.

Objections to this work have been voiced (Goldanski *in* Avetisov et al. 1987). For one thing, these authors (as well as L. Orgel, personal communication) maintain that the electroweak advantage takes place by repetitive steps in these calculations and that there are $n = 10^{17}$ steps involved. The earth must therefore have contained $n^2 \approx (10^{17})^2 \approx 10^{34}$ chiral molecules to take proper account of the resulting fluctuations. This, these authors (Goldanski *in* Avetisov et al. 1987) find difficult to credit, and conclude that "the role of weak neutral currents in the origination of the biomolecular chirality should not be considered essential." Kondepudi (personal communication) agrees with this estimate (10^{34}) but maintains that "this number of molecules need not all be reacting at the same time. This is the total number that is fluxed through the nonequilibrium flow system in 1.5×10^4 years."

Without necessarily disagreeing with the work of Kondepudi and Nelson (particularly at high temperatures where the equations they use may be considered part of the renormalization group), we speculate on an alternative mechanism. The discussion in this note is mainly physical in character. The enhancement due to phase transitions is discussed in Appendices A–E.

I would like to treat this quantum mechanically as an equilibrium problem. The quantum mechanical formalism treats the phenomenon as a cooperative one where condensation aspects are emphasized and the transitions D to L are accomplished below a critical temperature T_c.

In general, when global cooperative and condensation phenomena do take place, low temperatures (or high densities, e.g., as for neutron stars) are necessary; now if T_c for the amino acid turns out to be very small, then it would be plausible that the origin space in which production of chiral amino acids takes place, was larger than the Earth. This problem is discussed below.

Section 4

A modern version of the phenomenon of condensation has been described by Leggett (1990) giving a uniform treatment of a superfluid (like liquid helium II), where T_0—the so-called degeneracy temperature—is $\approx 3°\text{K}$, as well as of metallic superconductivity for Cooper-paired electrons below a transition temperature T_c. The latter is more relevant for our purposes as we shall see below.

What is condensation?

> Imagine that you are on a mountain-top looking down at a distant city square. The crowd is milling around at random, and each individual is doing something different: Now suppose, however, that it is not market day but the day of a military parade, and the crowd is replaced by a battalion of well drilled soldiers. Every soldier is doing the same thing at the same time, and it is very much easier to see (or hear)

from a distance what that is. The physics analogy is that a normal system is like the market day crowd—every atom is doing something different—whereas in a Bose condensed system the atoms (or, more accurately, the fraction of them which is condensed at the temperature in question) are all forced to be in the same quantum state, and therefore resemble the well drilled soldiers: every atom must do exactly the same thing at the same time (Leggett 1990).

The analogy of the behavior of the crowd with racemicity and of the well-drilled soldiers with L amino acids is apt. The number of particles in any given energy level is fixed as a function of temperature, and as a result the total number of particles occupying the levels cannot be greater than some number $N_{max}(T)$ (which decreases along with T). At some temperature T_0 the quantity N_{max} becomes equal to the total number of particles in the system N, whereas below T_0 we have $N_{max} < N$. At such temperatures there are simply not enough quantum states available to accommodate all the particles. "The resolution of the problem is remarkably simple: below T_0 the system adjusts by taking all the particles which cannot be accommodated by the distribution formula and putting them in the single quantum state which has the lowest energy" (Leggett 1990). Because these surplus particles are a finite fraction of the whole (in fact, at zero temperature all of them), we reach the remarkable result that a macroscopic number of particles (of order N, which typically is of order 10^{23}) occupy a single quantum state. This phenomenon is known as Bose condensation.

Section 5

The necessary conditions for condensation have been studied by L. Landau (Landau et al. 1980). Landau distinguished between Bose superfluids and Fermi superfluids (even for Fermi superfluids like superconductors the "superfluidity" is produced by bosonic condensates like those for Cooper pairs). [A condensate is the constant part of a spin zero field $\langle\varphi\rangle$ that can arise for some theories (provided the potential for φ is like an inverted Mexican hat). This is achievable for parts of fields that carry zero frequency (zero energy and zero momentum).]

"A Fermi gas with attraction between the particles must have the property of superfluidity . . . howsoever weak the attraction is" (Abrikosov 1987). To clarify this point further, consider the following situation (Goodstein 1985):

Imagine two people on an old sagging, nonlinear mattress. They tend to roll toward the middle, even if they don't like each other. That is, there is an attractive interaction. The cause of this interaction . . . is that the people create distortions in the mattress, and the distortions try to merge. The electrons in the metal do not stand still but rather zip through the lattice at something like the Fermi velocity. The ions are attracted to the electrons but, owing to their large mass, move very slowly compared to the much lighter electrons. By the time the ions respond the electron is long gone, but it has, in effect, left behind a trail of positive charge, which is the lattice distortion we mentioned above. Another electron, transversing the same path, would find that its way has been prepared with the positive charge that it finds so attractive. We can imagine that the first electron created a phonon, which the second happily absorbs. Notice that the interaction is strongest if the two electrons traverse exactly the same path—that is, if they have, say, equal and opposite momenta.

Section 6

For $T = T_c$ the phase transition is present with all its aspects of suddenness. For $T < T_c$ the "superfluidity" persists, reaching its climax at $T = 0$. The question arises as to what happens when $T > T_c$. To answer this question, let us consider cooperative phenomena. The main features of cooperative phenomena can be illustrated by considering with Atkins (1959), the semiclassical treatment of ferromagnetism. At 0°K the electron spins are aligned parallel to one another to give a resultant magnetization even in the absence of an external magnetic field. However, at a finite temperature thermal agitation is able to turn over some of the spins and the average magnetic moment in the direction of magnetization is thereby decreased. "As soon as this process starts, an electron chosen at random is likely to have neighbors pointing against the direction of magnetization as well as with it and this reduces the energy needed to reverse the spin of the electron, so that as the temperature increases and more spins are turned over, it becomes increasingly easier to turn over the remaining spins and the disordering process develops with ever-increasing rapidity. At the Curie point the disordering is eventually complete and the spins point equally in both directions" so that only a 50–50 racemic state survives.

The results of this section will be used in Appendix B to obtain one estimate of T_c for most amino acids.

Section 7

How important are the effects of the parity-violating weak interactions for specific amino acids? This question may be answered by setting $(1 - 4 \sin^2\theta) \approx 0$; the resulting elegant expression for the Hamiltonian has led to molecular conformation-dependent values for parity violation, which for Ala, Val, Ser, and Asp have been calculated to give (Mason and Tranter 1984; Tranter and MacDermott 1989) -3.0, -6.2, -2.3, and -4.8, in units of 10^{-19} eV. These authors express the energy values in atomic units (1 atomic unit = 1 Hartree = 27.2 eV) corresponding to the L configuration of the aqueous

zwitterionic amino acids. The simplest amino acid, glycine (residue R = H), is not resolvable into optical isomers. On account of the further symmetry implied by R = H, parity violation, when averaged over a complete rotation of the carboxylate group through all configuration angles, turns out to be zero, in agreement with experimental results. For the sugar glyceraldehyde the value is $+1.8 \times 10^{-19}$ eV corresponding to its D configuration [see section 1, Appendix E].

The classical chemist has hitherto used the electromagnetic force as the only fundamental force that can produce chemical effects. He has not considered the electroweak force, and in particular its Z^0 component, because the effects due to Z^0 are supposedly very small at low temperatures. I shall now show that the electroweak interactions (for which there is nothing comparable in classical chemistry) can give rise to phase transitions.

The superconductivity phase transition can be attributed to the attractive force due to the parity nonconserving effective interaction produced by the spin-0 part of Z^0, which is itself caused by spontaneous symmetry breaking associated with the condensate $\langle \varphi \rangle \neq 0$.

Like all phase transitions, this one will have a behavior like $(T_c - T)$, $T \to T_c$. This makes the amplitude (or derivatives thereof) an infinite quantity.

To compute T_c exactly is difficult (for one thing, on account of the incompleteness of the standard model of elementary particles), but T_c can be measured without much difficulty. In the sequel it is assumed that T_c does exist and that it has been measured for each amino acid.

Section 8

Because the ambient temperature of the Earth's surface is $\approx$300°K (the maximum temperature on the surface being $\approx$350°K), if $T_c \ll$ 300°K, the present formalism may not apply to the Earth. The fact that the prebiotic temperatures may indeed be less than 300°K was the content of a paper by Sanchez et al. (1966); these authors claim that tetramer formation in 0.01 M HCN is accelerated by lowering the temperature from 300°K to 250°K. The authors conclude that "we may have to replace the usual picture of a warm dilute prebiotic medium with one more cold and much more concentrated at least for some syntheses." Thus in the language of the present note, the data of Sanchez et al. may perhaps be interpreted as advocating $T_c \lesssim$ 250°K. On the other hand, K.D. Kondepudi (personal communication) has suggested that unless $T_c \lesssim$ 2.7°K—the ambient universal temperature—the entire interstellar space could be optically active provided, of course, the density of amino acids is not too dilute over most of space.

In this case, there could be several possible scenarios:

1) Presolar contributions may be necessary to get low enough temperatures.

2) Major contributions come from the more distant and the cooler parts of the solar system with ambient temperatures less than T_c. In this context it is good to remember that the maximum surface temperatures have been estimated at 135°K for Jupiter, 120°K for Saturn, 85°K for Uranus, 55°K for Neptune, and 20°K for Pluto (see section 2, Appendix E).

3) If T_c is considerably less than 300°K, then one could entertain the (somewhat odd) notion that the Earth only acted as a junction place where L amino acids came together with D sugars and nucleotides for the replication phenomena to get started for the biotic stage mentioned in Section 2 to be implemented.

In order for the biotic alternative to prevail, it is necessary to invent a mechanism to deliver organic molecules to the Earth from the cooler locations mentioned above. This was considered by Orò (Orò 1961; Orò et al. 1990a), and in greater detail recently by Chyba et al. (1990). The latter authors conclude that:

1) "The Earth did accrete prebiotic organic molecules important for the origins of life from impacts of carbonaceous asteroids and comets during the period of heavy bombardment 4.5×10^9 to 3.8×10^9 years ago" (cf. the Chiron comet?).

2) For plausibly dense (10-bar carbon dioxide) early atmospheres, these authors find that 4.5×10^9 years ago the Earth was accreting intact cometary organics at a rate of at least $\approx 10^6$ to 10^7 kg per year—a flux that thereafter declined with a half-life of $\sim 10^8$ years. These results may be placed in context by comparison with terrestrial oceanic and total biomasses, $\sim 3 \times 10^{12}$ kg and $\sim 6 \times 10^{14}$ kg, respectively.

3) Contrary to the general impression, alanine molecules could withstand temperatures as high as $\approx$700°K for 1 s whereas other amino acids could withstand temperatures in the range of 600–800°K for a like period of time and remain stable and intact upon impact with the Earth.

It is tempting to assume further that this material ($\approx 10^6$ to 10^7 kg) maintained its chirality at delivery (particularly if T_c had a small value near zero). This follows from the formula

$$-\frac{\partial \Delta}{\partial T} \approx \frac{3.06}{2} \times \sqrt{\frac{T_c}{T_c - T}}$$

(see Appendix B.2) if the formula applies to the case

of amino acids. At this juncture life could have started and the chirality perpetuated in accordance with the biotic picture mentioned in Section 2.

Section 9

Is the laboratory testing of the central hypothesis of this note feasible? This can be tested by taking a racemic mixture of crystalline L and D amino acids.

The crystalline conformations of L and D amino acids do not differ much from the zwitterionic forms except for the absence of water molecules inside the lattice cell. Barring for alanine, there are no definite calculations which show that it is indeed the L configuration that prevails for these crystals. However this may not affect the laboratory testing of the hypothesis of this paper because lowering of the temperature below the transition temperature would eventually convert all amino acids into that particular configuration that only partially predominated before. Conversely, a raising of temperature through T_c would show that a pure configuration below T_c can be converted into a racemic mixture for $T > T_c$.

To conclude, note that ideally one should be able to compute the values of T_c when electroweak theory is fully worked out. The numerical value of this quantity could lie anywhere around 2.7°K (the ambient temperature of outer space) or beyond 350°K but below the dissociation temperatures.

One could contemplate reaching 0°K, but this would have to be strictly local. The best way to determine T_c for a given amino acid is at present by experimentation. The analogy of the "superfluidity" exhibited by amino acids and sugars is to "superfluidity" in superconductors and not to the liquid helium. In the case of superconductivity, one has to apply an external magnetic field and look for the Meissner effect to determine T_c. Likewise, the "superfluidity" of amino acids (or sugars) is measured by shining on these external light sources (see section 3, Appendix E).

One direct way to test for evidence of the hypothesis (regarding the existence of such a phase transition) is to lower the temperature while measuring optical activity when polarized light is shined upon a racemic mixture of a particular amino acid. If the polarization vector gets rotated, one may be sure that the appropriate phase transition has taken place.

The process itself could perhaps be detected by optical rotatory dispersion (ORD) for circular dichroism (CD) (see section 4, Appendix E). An alternative means of detecting the process may be by measuring differences of specific heats and looking for anomalies in the curve $C = \gamma T + \beta T^3 + \ldots$ like what has been done (see Appendix D) for the non-amino acids like melanins and tumor melanosomes (Mitzutani et al. 1976).

Section 10

I have shown that chirality may provide a boundary condition for theories of the origin of life and that Z^0 interactions—as well as what comes beyond the standard model of fundamental interactions (CP violation, for example)—ought to play a central role in this story. In this context, the following quotation from L. Pasteur (who did not even know of Z^0 particles) is perhaps prophetic: "Life as manifested to us is a function of the asymmetry of the Universe and of the consequences of this fact. The Universe is asymmetrical. Life is dominated by asymmetrical actions. I can even imagine that all living species are primordially in their structure, in their external forms a function of cosmic asymmetry" (Pasteur 1860).

Acknowledgments. The author thanks Drs. L. Orgel, S.F. Mason, and J. Orò for kindly making available their recent work. Appreciation is expressed to Drs. F. Hoyle, K.D. Kondepudi, A.J. MacDermott, C. Ponnamperuma, P.G.H. Sandars, G.E. Tranter, C, Wickramasinghe, and Yu Lu for commenting on the early drafts of this manuscript. Particular thanks are due to Dr. L.N. Johnson of the Laboratory of Molecular Biophysics at Oxford and Drs. J. Chela-Flores and J. Strathdee at the International Center for Theoretical Physics, Trieste, for collaboration and wise counseling at every stage of this paper. The author also thanks Ms. S. Radosic of the Publications Section of the I.C.T.P. for patiently typing and retyping the manuscript from dictation.

References

Abrikosov AA (1987) Fundamentals of the theory of metals. Elsevier, North Holland, pp 345–347

Arnold P, McLerran L (1987) Sphalerons, small fluctuations and baryon-number violation in electroweak theory. Phys Rev D36:581–595

Atkins KR (1959) Liquid helium. Cambridge University Press, New York, p 42

Avetisov VA, Kus'min VV, Anikin SA (1987) Sensitivity of chemical chiral systems to weak asymmetric factors. Chem Phys 112:179–187

Chela-Flores J (1985) Evolution as a collective phenomenon. J Theor Biol 117:107–118

Chyba CF, Thomas PJ, Brookshaw L, Sagan C (1990) Cometary delivery of organic molecules to the early Earth. Science 249: 366–373

Cronin JR (1989) Origin of organic compounds in carbonaceous chrondrites. Adv Space Res 9:54–64

Cruikshank DP (1989) Dark surfaces of asteroids and comets: evidence for macromolecular carbon compounds. Adv Space Res 9:65–71

Delbrück M (1963) In commemoration of the 50th anniversary of Niels Bohr's papers on atomic constitution, Session on Cosmos and Life, Institute for Theoretical Physics, Copenhagen, pp 41–67

Engel MH, Macko SA, Silfer JA (1990) Carbon isotope composition of individual amino acids in the Murchinson meteorite. Nature 348:47–49

Goodstein DL (1985) States of matter. Dover Publications, New York, pp 386–391

Hanel R, Conrath B, Flasar M, Kunde V, Lowman P, Maguire W, Pearl J, Pirraglia J, Samuelson R, Gautier D, Gierasch P, Kumar S, Ponnamperuma C (1979a) Infrared observations of the Jovian system from Voyager 1. Science 204:972–976

Hanel R, Conrath B, Flasar M, Herath L, Kinde V, Louman P, Maguire W, Pearl J, Pirraglia J, Samuelson R, Gautier D, Gierasch P, Horn L, Kumar S, Ponnamperuma C (1979b) Infrared observations of the Jovian system from Voyager 2. Science 206:952–956

Harris MJ, Loving CE, Sandars PGH (1978) Atomic shielding and PNC optical rotation in bismuth. J Phys B11:L749–L753

Knervolden KA, Lawless J, Ponnamperuma C (1971) Nonprotein amino acids in the Murchinson meteorite. Proc Natl Acad Sci USA 68:486–490

Kondepudi KD, Nelson CW (1985) Weak neutral currents and the origin of biomolecular chirality. Nature 314:438–441

Kuzmin V, Rubakov V, Shaposhnikov M (1985) On the anomalous electroweak baryon-number non-conservation in the early universe. Phys Lett 155B:36–42

Landau LD, Lifschitz EM, Pitaevskii LP (1980) Statistical physics, part II, vol 9. Pergamon, New York

Leggett A (1990) In: Davies P (ed) The new physics. Cambridge University Press, New York, p 276

Mason SF, Tranter GE (1984) The parity-violating energy difference between enantiomeric molecules. Mol Phys 53:1091–1111

Mitten S (1977) The Cambrige encyclopædia of astronomy. Trewin Copplestone, London, p 392

Mizutani U, Massalski TB, McGiness JE, Corry PM (1976) Low temperature specific heat anomalies in melanins and tumor melanosomes. Nature 259:505–507

Orò J (1961) Comets and the formation of biochemical compounds on the primitive earth. Nature 190:389–390

Orò J, Mills T (1989) Chemical evolution of primitive solar system bodies. Adv Space Res 9:105–120

Orò J, Miller SL, Lazcano A (1990a) The origin and early evolution of life on Earth. Annu Rev Earth Planet Sci 18: 317–356

Orò J, Squyres SW, Reynolds RT, Mills TM (1990b) Europa: prospects for an ocean and exobiological implications. NASA Spec Publ (in press)

Pasteur L (1860) Researches on the molecular asymmetry of natural organic products. Alembic Club Reprint, no. 14, London

Ponnamperuma C, Molten P (1973) Prospect of life on Jupiter. Space Life Sciences 4:32–44

Randjbar-Daemi S, Salam A, Strathdee J (1990) Chern–Simons superconductivity at finite temperature. Nucl Phys B340:403–447

Ringwald A (1990) High energy breakdown of perturbation theory in the electroweak instanton sector. Nucl Phys B330: 1–18

Sakita B (1985) Quantum theory of many-variable systems and fields. World Scientific, Singapore

Sanchez R, Ferris J, Orgel LE (1966) Conditions for purine synthesis: did prebiotic synthesis occur at low temperatures? Science 153:72–73

Soderblom LA, Kieffer SW, Beckes TL, Brown RH, Cook AF, Hansen CJ, Johnson TV, Kirk RL, Schoemaker EM (1990) Triton's geyser-like plumes: discovery and basic characterization. Science 250:410–415

Tranter GE, MacDermott AJ (1989) Electroweak bioenantioselection. Croatica Chem Acta 62(2A):165–187

Walker DC (ed) (1979) Origins of optical activity in nature. Elsevier, New York, p vii

Wyckoff RWG (1966) Crystal structures, vol 5, ed. 2. Interscience, London 5, pp 725–727

Received August 30, 1990/Revised March 7, 1991

Appendix A

Section A.1

To implement these ideas for L amino acids we start with the Z^0 interaction

$$L_{\text{int}} = \frac{e}{\sin\theta\cos\theta}\left[(T_{3L} - \sin^2\theta\, J_{\text{em}})_\mu Z^0_\mu\right] \qquad \text{(A.1)}$$

(units $\hbar = c = 1$). Here J_{em} is the electromagnetic current, T_{3L} is the left-handed third component of the weak isospin consisting of the (anomaly-free) combination of the proton–neutron (p, n) and neutrino–electron (ν_e, e) left-handed doublets (of weak isospin), i.e., $T_{3L} = \bar{\psi}\gamma_\mu\left(\frac{1+\gamma^5}{2}\right)\tau_3\psi$ where $\tau_3 = \frac{1}{2}(1, -1)$ where ψ stands either for the doublet of protons (p) and neutrons (n), or for the doublet of neutrinos (ν_e) and electrons (e). Proton and neutron (composite) fields have been used in preference to the elementary quarks, as this illustrates the point that for our present calculations (up to certain energies and temperatures), it is not important to know the elementary entities themselves. In addition there is the Higgs field φ, with a known nonzero expectation value $\langle\varphi\rangle$, which is needed to give masses to the protons, neutrons, and electrons as well as the Z^0 particle. Choosing the value of $\langle\varphi\rangle \neq 0$ is the same as spontaneous symmetry breaking. The Higgs particles on account of their nonzero expectation value ($\langle\varphi\rangle \neq 0$) can act as perfect condensates at zero temperatures. In terms of this quantity ($\langle\varphi\rangle = 250$ GeV) the electron mass m_e turns out to be a very tiny number ($\approx 2 \times 10^{-6} \times \langle\varphi\rangle$). [It is perhaps worth remarking that this number is very large for the top quark if its mass is in excess of 100 GeV. Thus the terms that give this, look like $b\varphi t\bar{t}$ where $b \geq \frac{2}{3}$. Some physicists, like Y. Nambu, take this as the defining property of the field ϕ, i.e., ϕ is considered as a $t\bar{t}$ composite.]

Let $(1 - 4\sin^2\theta) \approx \frac{1}{13}$ with the present empirical value of the parameter $\sin^2\theta \approx 0.231$. Unlike some authors (Harris et al. 1978; Mason and Tranter 1984; Tranter and MacDermott 1989) let us not assume this quantity to equal zero. Neglecting neutrinos (and Higgs), the right-hand side of (A.1) can be written in the form:

$$\frac{e}{4\sin\theta\cos\theta}(V_\mu - a_\mu)Z_\mu \qquad \text{(A.2)}$$

Here

$$V_\mu = (1 - 4\sin^2\theta)J_{\text{em}\mu} + (-\bar{n}\gamma_\mu n) \qquad \text{(A.3)}$$

and

$$-a_\mu = (\bar{p}\gamma_\mu\gamma_5 p - \bar{n}\gamma_\mu\gamma_5 n) - (\bar{e}\gamma_\mu\gamma_5 e) \qquad \text{(A.4)}$$

To this must be added the purely electromagnetic terms (which are parity conserving)

$$eJ_{\text{em}\mu} \times A_\mu \qquad \text{(A.5)}$$

Section A.2

In order to address the question of the existence of a finite critical temperature T_c for the phase transition into a condensed mode, we recall that our picture of amino acids is as follows:

Amino acids consist of a backbone made up of atoms of

carbon (2 in number, one of which C_α acts as the center for the mirror transformations), oxygen (2), nitrogen (1) and hydrogen (4), plus a residue that may consist of hydrogens (up to 15), carbons (up to 8), nitrogens (up to 3), oxygens (up to 2), and sulfur (up to 1). The electrons interact with protons and neutrons essentially (in fact with the quarks contained inside the nucleons) at the location of these objects; they in particular interact with neutrons.

The amino acid gets its L or D configurations through mirror arrangements of the atoms contained in the backbone structure. It has been shown that L or D configurations are correlated with left or right chirality (Mason and Tranter 1984).

The state of a given amino acid as composite of p, n, e, and ν will be shown in the sequel. It must be remembered that the amino acids are not metallic compounds. Thus there is no concept of Fermi energy nor of wandering electrons for them. The electrons from the inner shells interact with the nucleons, in particular with the neutrons and quarks inside them.

Appendix B

Section B.1

Consider the pseudoscalar terms in the effective interactions, which, after integrating out the Z^0 field, is proportional to $V_\mu \times A_\mu + A_\mu \times V_\mu \approx \bar{\psi}\gamma_\mu\psi\bar{\psi}'\gamma_\mu\gamma_5\psi' + \ldots$. A part of these terms is proportional to $(1 - 4\sin^2\theta)$. Such terms have been set equal to zero by some authors (Harris et al. 1978; Mason and Tranter 1984; Tranter and MacDermott 1989). I shall not make this approximation. These terms contain 4-Fermi interaction of electrons that are proportional to the pseudoscalar quantity $\sigma\cdot p$ in the nonrelativistic approximation, where only large components are kept. Such terms are proportional to $(1 - 4\sin^2\theta)[(e^+e \times e^+\frac{\sigma\cdot p}{m_e}e) + (e^+e \times n^+\frac{\sigma\cdot p}{m_n}n)] + n^+n \times e^+\frac{\sigma\cdot p}{m_e}e$. The important point for our purposes is that this part of the parity-violating sector of e–e coupling is negative for half the states of the system, i.e., for states that have the opposite eigenvalue for the operator $\sigma\cdot p$ before and after the interaction—and positive for the remaining half of the states.

I concentrate on the attractive set among the parity-violating terms. Because of the absence of such terms in the parity-conserving (in fact repulsive) Coulomb force, this would allow us to develop the analogy for metals within the theory of superconductivity. The situation is analogous to that of BCS theory where the 4-Fermi effective interaction of the electrons is attractive. According to Landau's criterion, the electron fluid must therefore exhibit "superfluidity" (Abrikosov 1987).

Section B.2

The ideas and results from BCS theory, which may be of interest to us in developing our analogy, are next emphasized.

The condensate wave function for a metallic superconductor is the "gap" function $\Delta(0) = \omega_D\exp\left(\frac{-2}{gv}\right)$. Here ω_D is the Debye cutoff $\approx 10^2K$ for most metals, p_F stands for the Fermi momentum, ω_F is the corrresponding Fermi energy, and g is the effective 4-Fermi coupling parameter for electrons, which in general is given by $v = \frac{p_F m}{\pi^2}$, whereas $gv \approx 1$ represents an approximation. Thus the expression for the exponent may be taken to be of order unity (as $gv \approx 1$). I shall continue to make this approximation. Empirically, one finds that $\omega_D \approx (10^{-3}$–$10^{-4})\omega_F$: ["The fact that the interaction spreads over an energy interval $\Delta\omega_D$ implies, according to quantum mechanics, that it is retarded, or in other words, it operates during a finite time interval $\Delta t \approx (\Delta\omega_D)^{-1}$" (Abrikosov 1987). This implies that for the electron–phonon interaction, the time interval is given by approximately 10^{-13} s for $T_c = 10$°K for the case of niobium (superconductivity) and 10^3 s for the case of amino acids if $\omega_D \approx 3 \times 10^{-19}$ eV.]

Using the methods of Gorkov and Sakita for the nonrelativistic electron case, an equivalent Landau–Ginzburg equation for the BCS theory is written down. This gives the following results for the superconductivity case:

1) T_c, the critical temperature is given by $T_c \approx (1.76)^{-1} \times \Delta(0)$.

2) The dependence of $\Delta(T)$ on temperature is approximately given by $\Delta(T) \approx \pi\left(\frac{8}{7\zeta(3)}\right)^{1/2}[T_c(T_c - T)]^{1/2} = 3.06[T_c(T_c - T)]^{1/2}$ for $T < T_c$; here $\zeta(x)$ is the Riemann zeta function: $\zeta(x) = \sum_{n=1}^{\infty} n^{-x}$. $\Delta(T)$ decreases with increasing temperature. Its derivative with respect to T becomes infinite at $T \approx T_c$ (Abrikosov 1987).

3) $C_s(T_c) = C_n(T_c) + \frac{4}{7\zeta(3)}\frac{p_F m}{\hbar^3}T_c$. Here C_s corresponds to specific heat for the superconducting phase, whereas C_n denotes the corresponding quantity for the normal case. From this, the result that $[C_s(T_c) - C_n(T_c)]$ is proportional to T_c is obtained. The same formula seems to apply for some of the organic materials as shown by Mizutani et al. (1976).

4) The expression 2 above, for $\Delta(T)$, holds for $T < T_c$. What happens in general for $T > T_c$? Goodstein states that for the (analogous) case of magnetic susceptibility, this expression is of the form, $f_+ \times \epsilon^{-\gamma}$ for $T > T_c$, provided the susceptibility for $T < T_c$ has the form $f_-(-\epsilon)^{-\gamma'}$, where $\epsilon = \frac{T - T_c}{T_c}$.

One consequence of scaling laws of physics is that $\gamma = \gamma'$ (Goodstein 1985, p. 481).

If these results can be carried over to the amino acid case, a racemic mixture starts forming for $T > T_c$ completing, in general, the process to a 50–50 mixture for $(f_+ + f_-)T_c$. Because the melting point (m.p.) represents the dissociation limit for the amino acids, one expects that $(f_+ + f_-)T_c$ is less than $T_{\text{m.p.}}$. Mr. P. Agbedjro, to whom thanks are due, has compiled from the 1988–1989 "Handbook of Chemistry and Physics," the following table of the melting points of the amino acids: Ala 568°K, Arg 537°K, Cys 533°K, Glu 497°K, Gly 535°K, His 560°K, Pro 511°K, Ile 557°K, Leu 566°K, Lys 497°K, Met 554°K, Phe 557°K, Pro 493°K, Ser 519°K, Try 563°K, Tyr 615°K, and Val 571°K. These numbers uniformly lie between 500°K and 600°K except for Tyr, which is ≈615°K.

This could give 250–300°K if $f_+ = f_- = 1$ [compare the work of Sanchez et al. (1966), cf. Section 8] for one estimate of T_c. However this estimate could vary between a very wide range of values if $f_+ \neq f_- \neq 1$.

5) The present BCS theory applies to the low temperature case. How are the ideas of this theory utilized for the case of high T_c? Our interest is in the critical temperatures for amino acids that may well be in excess of 350°K. Such a model beyond BCS might, for example, follow more the analogy of high T_c superconductivity where present experiments take T_c to values as high as 125°K. There is, as yet, no accepted theory (Randjbar-Daemi et al. 1990) of these high T_c superconductors and it is not clear what, if anything, is the analogue of Cooper pairs.

Two of the mechanisms suggested that high T_c superconductivity may possibly be of relevance to the case of "superfluidity" for amino acids. These are: (1) The use of the electrons of inner shells of atoms that are retained in the lattice ions. ω_D could be as high as $\approx 10^{20}$K; and (2) an attempt to find a transmitting system of electrons with high polarizability: molecular organic crystals, polymers, for example C_{xn} (where C stands for carbon and x represents a multiple repetition of the group). Unfortunately, as far as high T_c superconductivity is concerned, neither

of these two suggestions have proven very effective in increasing T_c to higher values.

Appendix C

To proceed for the case of amino acids from these expressions is not that easy: the difficulties in carrying this program through are recounted below.

Section C.1

For the amino acids, let us take an expression for T_c that is similar in form to the expression for superconductivity $T_c = \Delta E \exp(-1/\lambda)$. "$\Delta E$ is the energy difference between the states A and A^*. A is the ground state, A^* is the excited state of the . . . system (it can easily be shown that such an interaction is necessarily an attraction if it is not strong)." λ depends on the interaction of electrons for the case of superconductivity, and ΔE is related to ω_D, and $\lambda \propto gv$. Let us designate $\exp(-1/\lambda)$ as the probability factor for the transition A to A^* to take place. The major problem will be to secure $gv \approx 1$ so that this probability factor is not too small.

1) A nonperturbative calculation like the one used by Sakita (1985) is needed for the case of superconductivity but for attractive parity-violating interactions. The difficulty in this case lies in choosing what the analogy is for the Fermi energy.

2) The electrons in the inner shells of atoms are probing much more deeply into the quark structure of the nucleon. In Appendix B such electrons were observed to give rise to 10^{20}K in the superconductivity case. Chemical phenomena are probing deeply into the centers of nucleons (and the quarks within them), the energies involved being much higher than one is used to.

3) $gv \approx 1$ may be difficult to achieve. Because $\lambda \propto gv$ the fact that $gv \neq 1$, in general, may mean a diminution of probability represented by $\exp(-1/\lambda)$.

Section C.2

There is another possibility, that is, to consider the relativistic term $\bar{n}\gamma_\mu n \times \bar{e}\gamma_\mu\gamma_5 e$. This term does not have the factor $(1 - 4 \sin^2\theta)$ in front of it.

Could we utilize this term to invent a condensate model of $\bar{e}$–n pairing like the Cooper pairing? The answer is clearly yes, as can be seen by the fact that an interchange of particles 2 and 4 by a Fierz reshuffle can be made, so that the term reads $\bar{n}e \cdot \bar{e}\gamma_5 n$. Both factors contained here are scalars and a field φ' can be invented that could have the effective coupling $[\varphi' \times (\bar{n}e + \bar{e}\gamma_5 n)]$. From the calculations previously made (Mason and Tranter 1984; Tranter and MacDermott 1989), for this particular case if $T_c \approx \omega_D \exp(-2/g'v')$ and if $g'v' \approx 1$, ω_D can be taken as 3×10^{-16} eV.

What field does φ correspond to? Clearly an $\bar{e}$–n pairing (with $\Delta B \neq 0$, $\Delta L \neq 0$) brings back the memories of proton decays and baryon number violation. The question would then be, is this triggered off by a grand unification mass $\approx 10^{14}$–10^{15}GeV or does it rely upon the work of Rubakov and his collaborators (Kuzmin et al. 1985; Arnold and McLerran 1987; Ringwald 1990), which ascribes baryon violation to the standard model with $\langle\varphi\rangle = 250$ GeV? In other words is $\langle\varphi\rangle = \langle\varphi\rangle'$? I would tend to favor this, but this does bring us to the modern unresolved controversies that are the substance of present-day theoretical particle physics.

There is a further uncertainty that comes about because of the uncertainty in physics of the standard model. This theory (with Z^0 particles) violates P and C but conserves CP. If parity violation P is accompanied by CP ($\cong$T) violation (and this seems to be the case from the K* system), this is accomplished by writing down a mass matrix of the Cabibbo–Kobayashi–Maskawa type. Such a mass matrix has been shown to involve necessarily a phase factor if there are three generations. A different procedure would be to build this violation as a consequence of extra terms as in Weinberg's new theory that uses gluons. It could also be built in by postulating the existence of invisible axions, where the appropriate mass terms may be greater than 10^{12} GeV, or by enhancing the symmetry of theory to left–right symmetry, or of the idea above of considering the condensate $\langle\varphi'\rangle \neq 0$ of spin-0 particle, which has interactions with $\bar{n}e$ and $e\gamma_5 n$. This may mean the addition of extra terms to the standard model, and more terms added for T_c like the ones already included.

Section C.3

The value 250°K obtained in Appendix B.4, if correct, could provide one of the better possibilities physically. It would mean that the Earth had a reasonable chance of being the site of prebiotic L amino acid production. [The very best possibility is, of course, $T_c \geq 350$°K. This would mean that the Earth suffices for the production site of amino acids, in general. This is a possibility that should not be ignored if laboratory tests are made for amino acids in accordance with what is suggested in Section 9.]

All that may reasonably be inferred at this stage is the existence of a nonzero T_c due to the attractive forces mentioned before.

Be that as it may, consider the case of 4-Fermi interactions with electrons. Because this term comes together with the factor $(1 - 4 \sin^2\theta)$, the form to be expected is of the type

$$T_c \approx \frac{\langle\varphi\rangle}{10^3} \exp[-2/gv(1 - 4 \sin^2\theta)] \approx 2.5 \times 10^{20}\text{K} \qquad \text{(C.1)}$$

Here ΔE is taken to be $\approx\langle\varphi\rangle$ multiplied by the old familiar factor of 10^{-3}, while gv is still taken ≈ 1. The exponential factor gives $\exp -26 \approx 10^{-10}$, so that altogether one has $10^{-13} \times \langle\varphi\rangle$.*

Because the only quantity that has the right transformation character (for such parity-violating interactions) is Z^0_μ, or its longitudinal part where $\partial Z_\mu/\partial x_\mu \approx M\varphi$, therefore 250 GeV must somehow play a role of in the formulae that have been set up.

To emphasize the arbitrariness in the calculation so far shown, it is worth remarking that T_c comes out to be $T_c = \omega_D(\exp - 2/(1 - 4 \sin^2\theta) = 5$°K if one takes $\omega_D \approx m_e \approx 2 \times \frac{\langle\varphi\rangle}{10^6}$.

Appendix D

After this paper was completed, Dr. J. Chela-Flores provided the following quotation to be inserted into the text. I gladly do so:

> Perhaps the earliest suggestion of the possible occurrence of condensation in biology was made almost three decades ago by Delbrück, who was concerned as to "whether or not something very peculiar from the quantum mechanical point of view, like superconductivity or superfluid helium, will come up. If strange cooperative phenomena can happen at room temperature in very special molecules . . . , then certainly life will have discovered this (Delbrück 1963). Within the context of the origin of life, condensation was conjectured to occur in the earliest riboorganism (Chela-Flores 1985)

* This is such an important point that I shall discuss it further in a separate note to be published in a physics journal. The top quark is necessary to get $gv = 1$. The reason for this is that g contains m_t^2, which is of the order of m_W^2.

under the effect of low temperatures estimated to be approximately 160°K.

Regarding the laboratory tests mentioned in the text, I thank Dr. A.J. MacDermott for sharpening the suggestion regarding the polarization measurements, and Dr. J. Chela-Flores for the suggestion regarding the use of specific heats.

Appendix E consists of four clarifications that occur in places in the text.

Appendix E

1) It may be noted in passing that if $(1 - 4 \sin^2 \theta) \approx 0$ uniformly, one is taking the contribution of the neutrons inside the nuclei into account, but not of the protons, except when deuterons substitute for protons. Also note that the isotope C^{13} should give a different contribution than C^{12}. This is the peculiar hallmark of Z^0 interactions. Such behavior has been attributed in the past to presolar cosmic abundances. For example, "the discovery that amino acids from the Murchison meteorite are as a group highly enriched in deuterium ($\delta D = 1370\%$) strongly suggests that the amino acids or their percursors were formed at low temperatures in interstellar clouds" (Chyba et al. 1990). This pathway reportedly supports the hypothesis of a direct relationship between organic-rich interstellar grains, comets, dark asteroids, and carbonaceous chondrites (Cruikshank 1989). It is clearly important to get the precise ratios of D/H, tritium/hydrogen, as well as C^{13}/C^{12} separately in order to distinguish the effects of Z^0 from the contributions due to abundances in the early universe. This has been done for the Murchison meteorite (Engel et al. 1990); these authors conclude that optically active materials were present in the early solar system before life began.

2) Such extraterrestriality had been anticipated on different grounds. Some 40 years ago, Stanley Miller performed fundamental experiments demonstrating that the action of electrical discharges on a mixture of water vapor, methane, and ammonia could produce reasonable yields of simple amino acids (glycine and alanine). Subsequent experiments along the same lines by Orgel, Ponnamperuma, and others have confirmed Miller's findings. The Miller–Orgel–Ponnamperuma "experiments simulated an environment rather similar to the lower atmospheric regions of Jupiter, where it is known that there are violent thunderstorms. Jupiter (or one of its satellites like Europa) may well be the best candidate in the solar system for rudimentary extraterrestrial life!" (Orò 1961; Ponnamperuma and Molten 1973; Mitton 1977; Hanel et al. 1979a; Orò and Mills 1989; Orò et al. 1990b). It is also conceivable that chemical evolution and synthesis of biochemical compounds have occurred and are occurring now in Titan—the largest satellite of Saturn. Titan has a reducing atmosphere—a desirable feature. This may explain in part, "some of the darker spots observed in Europa's outer surface and more recently in Triton, the remarkable satellite of Neptune" (Orò et al. 1990b). [Triton's surface temperature apparently is $\approx 38 \pm 4$°K (Soderblom et al. 1990).]

3) The difference between superconductivity and Bose superfluidity for He^4 lies in the fact that the Cooper pairs (which are bosons) are rather large objects ($\sim 10^{-4}$ cm) compared with the interparticle distance ($\sim 10^{-6}$ cm), i.e., there is a significant overlap between Cooper pairs.

4) It is conceivable that the experiments are best done where a heavy atom substitutes for one of the light atoms. The analogy could be with the Patterson phases in normal x-ray diffraction analyses where one uses atoms of gold, platinum, or mercury. I find that amino acid crystals have been made with nickel, iron, or copper, for example, crystals of silver glycine $AgOOC\text{-}CH_2\text{-}NH_2$ or copper DL-α-aminobutyrate $Cu(OOC \cdot C_2H_6NH_2)_2$ (Wyckoff 1966).

Physics Letters B 288 (1992) 153–160
North-Holland

PHYSICS LETTERS B

Chirality, phase transitions and their induction in amino acids

Abdus Salam
Department of Theoretical Physics, Imperial College, London SW7 2BZ, UK
and International Centre for Theoretical Physics, I-34014 Trieste, Italy

Received 11 May 1992

"Atoms such as carbon, oxygen, nitrogen and hydrogen, the major constituents of biological molecules, are less than 0.4 nm in diameter.... The behaviour of small molecules is a reflection of the intrinsic properties of the constituent atoms. Hence it might be expected that the behaviour of large macromolecules can be explained by a knowledge of atomic properties. Since organelles, whole cells and organisms are essentially macromolecular assemblies, it may be possible in time to derive an atomic theory of life" [A.R. Rees and M.J.E. Sternberg, From cells to atoms – An illustrated introduction to molecular biology (Blackwell, Oxford, 1984) p. 3]. It has been suggested that chirality among the twenty amino acids which make up the proteins may be a consequence of a phase transition which is analogous to that due to BCS superconductivity [A. Salam, J. Mol. Evol. 33 (1991) 105]. We explore these ideas in this paper and show, following Lee and Drell [I.H. Lee and S.D. Drell, in: Fermion masses in the standard model, M.A.B. Bég Memorial Volume, eds. A. Ali and P. Hoodbhoy (World Scientific, Singapore, 1991) p. 13], that a crucial form for the transition temperature T_c involves dynamical symmetry breaking. The t-quarks or supersymmetry (or something similar which ensures a heavy mass) appear to be essential if such mechanisms are to hold.

1. The T_c for BCS superconductivity for metals is of the form $\omega \exp[-2/g_{\text{eff}}\sigma(0)]$. We conjectured that a similar formula may hold for the case of amino acids chains. In the present paper we shall explore this further.

First, let us review the subject of BCS superconductivity. The best treatment, which I know of has been given by Sakita [1], following on the work of Ginsberg and Landau [2] and of Gorkov [3].

The idea is to start from the Feynman lagrangian methodology of writing down the BCS theory for the superconducting electronic system. One tries to write down the equivalent Ginzburg–Landau equation. From this equation is deduced the value of temperature T_c.

2. The BCS hamiltonian is given by

$$H = H_0 + H_1 ,$$

$$H_0 = \sum_{s=\uparrow\downarrow} \int \mathrm{d}x\, \psi_s^+(x) \left(\frac{2}{2m} (-\nabla) - \mu \right) \psi_s(x),$$

$$H_1 = -g_{\text{eff}} \int \mathrm{d}x\, \psi_\uparrow^+(x) \psi_\downarrow^+(x) \psi_\downarrow(x) \psi_\uparrow(x) , \tag{2.1}$$

where g_{eff} is an attractive coupling constant between spin up ($\uparrow$) and spin down ($\downarrow$) electrons and antielectrons is the chemical potential. *The sign of g_{eff} is part of the assumption* of the hamiltonian which signifies an attractive force between Cooper-paired systems of electrons consist of one of the particles being replaced by its antiparticle with a factor of two which appears in the mass term.

We can introduce electro-magnetic interaction in a gauge-invariant way by the minimal substitution

$$\nabla \to \nabla - \mathrm{i}eA ,$$

and treat the vector potential $A(x)$ as an external source. Then, in general, the hamiltonian equals

© 1992 Elsevier Science Publishers B.V.

$$H=\sum_{s=\uparrow,\downarrow}\int \mathrm{d}x\,\psi_s^+(x)\left(-\frac{1}{2m}(\nabla-\mathrm{i}eA)^2-\mu\right)\psi_s(x)-g_{\mathrm{eff}}\int \mathrm{d}x\,\psi_\uparrow^+(x)\psi_\downarrow^+(x)\psi_\downarrow(x)\psi_\uparrow(x)\,, \tag{2.2}$$

which is invariant under gauge transformations:

$$\psi_s(x)\to\exp[\mathrm{i}e\Lambda(x)]\psi_s(x),\quad A(x)\to A(x)+\nabla\Lambda(x)\,. \tag{2.3}$$

The partition function based on the appropriate lagrangian corresponding to the above hamiltonian is given by

$$Z=\int \mathrm{D}\psi\,\mathrm{D}\bar{\psi}\exp\left[-\int_0^\beta \mathrm{d}\tau\int_V \mathrm{d}x\,\bar{\psi}_s\left(\frac{\partial}{\partial\tau}-\frac{1}{2m}(\nabla-\mathrm{i}eA)^2-\mu\right)\psi_s\right],$$

where we have written a non-relativistic equation for the fermion (electron) ψ in the theory:

$$\exp\left(g_{\mathrm{eff}}\int_0^\beta \mathrm{d}\tau\int_V \mathrm{d}x\,\bar{\psi}_\uparrow(x)\bar{\psi}_\downarrow(x)\psi_\downarrow(x)\psi_\uparrow(x)\right). \tag{2.4}$$

Note that the sign before g_{eff} has changed to plus g_{eff} instead of minus g_{eff} in the hamiltonian formulation. This is because $\Delta H=-\Delta L$ for the case of potentials which do not contain time derivatives of the fields themselves. The four-fermion interaction can be expressed in terms of a complex auxiliary Higgs scalar field ϕ:

$$\frac{1}{C}\int \mathrm{D}\phi\,\mathrm{D}\phi^*\exp\left(+\kappa^2\int \mathrm{d}^4x\,\phi^*\phi+g_{\mathrm{eff}}^{1/2}\kappa\int \mathrm{d}^4x(\bar{\psi}_\uparrow\bar{\psi}_\downarrow\phi+\psi_\downarrow\psi_\uparrow\phi^*)\right), \tag{2.5}$$

where we have used the notation

$$\int \mathrm{d}^4x\equiv\int_0^\beta \mathrm{d}\tau\int_V \mathrm{d}^3x\,. \tag{2.6}$$

The constant C is given by

$$C=\int \mathrm{D}\phi\,\mathrm{D}\phi^*\exp\left(+\kappa^2\int \mathrm{d}^4x\,\phi^*\phi\right), \tag{2.7}$$

where κ is a constant with dimension of mass. We shall introduce a source for the ϕ field in the form

$$Z[j,j^*]=\frac{1}{C}\int \mathrm{D}\psi\,\mathrm{d}\bar{\psi}\,\mathrm{D}\phi\,\mathrm{D}\phi^*\exp\left[-\int \mathrm{d}^4x\,\bar{\psi}_s\left(\frac{\partial}{\partial\tau}-\frac{(\nabla-\mathrm{i}eA)^2}{2m}-\mu\right)\psi_s\right]$$

$$\times\exp\left(+\kappa^2\int \phi^*\phi\,\mathrm{d}^4x+g_{\mathrm{eff}}^{1/2}\kappa\int \mathrm{d}^4x(\bar{\psi}_\uparrow\bar{\psi}_\downarrow\phi+\psi_\downarrow\psi_\uparrow\phi^*+j^*\phi+\phi^*j)\right). \tag{2.8}$$

The partition function is given by

$$Z=Z[j,j^*]\,|_{j=j^*=0}\,. \tag{2.9}$$

The action which has appeared in (2.8) is $L\,\mathrm{d}\tau$ where

$$L=L_0+L_1\,, \tag{2.10}$$

$$L_0=\int \mathrm{d}x\left[\bar{\psi}_s\left(\frac{\partial}{\partial\tau}-\frac{\nabla^2}{2m}-\mu\right)\psi_s+\kappa^2\phi^*\phi\right], \tag{2.11}$$

Volume 288, number 1,2 PHYSICS LETTERS B 20 August 1992

$$L_1 = \int \mathrm{d}x \left(-\frac{ie}{2m} [\bar{\psi}_s \nabla \psi_s - (\nabla \bar{\psi}_s)\psi_s] A + \frac{e^2}{2m} \bar{\psi}_s \psi_s A^2 + g_{\mathrm{eff}}^{1/2} \kappa (\bar{\psi}_\uparrow \bar{\psi}_\downarrow \phi + \psi_\downarrow \psi_\uparrow \phi^*) \right). \tag{2.12}$$

Using the temperature dependent Fourier decomposition

$$\psi_s(x, \tau) = (\beta V)^{-1/2} \sum_k \psi_s(k, n) \exp[\mathrm{i}(k \cdot x - \xi_n \tau)] , \tag{2.13}$$

we obtain the free action

$$\int_0^\beta \mathrm{d}\tau L_0 = \sum_{k,n} [\bar{\psi}_s(k, n)(-\mathrm{i}\xi_n + \omega_k)\bar{\psi}_s(k, n) + \kappa^2 \bar{\phi}^*(k, n)\bar{\phi}(k, n)] . \tag{2.14}$$

Therefore, the propagator of the electron with four momentum (k, n) is given by

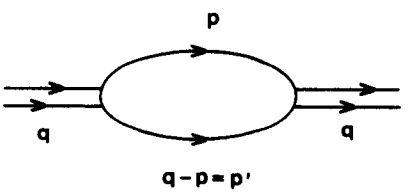
$$\frac{1}{\omega_k - \mathrm{i}\xi_n}, \quad \omega_k = \frac{k^2}{2m} - \mu, \quad \xi_n = \frac{\pi}{\beta}(2n+1) , \tag{2.15}$$

and the propagator of ϕ by

$$\frac{1}{\kappa^2} . \tag{2.16}$$

Notice that the propagator of ϕ to this order does not depend on four momentum, because the lagrangian (2.11) does not contain the kinetic energy term.

3. *Calculation of the loop diagram.* We shall compute one-loop diagram with two external ϕ-lines. The (temperature dependent) leading terms are



$$\overset{q_0 = 0}{=} \frac{g_{\mathrm{eff}} \kappa^2}{\beta V} \sum_{n,p} \frac{1}{(\mathrm{i}\xi_n - \omega_p)(-\mathrm{i}\xi_n - \omega_{p'})} = A + B\boldsymbol{q}^2 + \ldots , \tag{3.1}$$

where

$$A = \frac{g_{\mathrm{eff}} \kappa^2}{\beta V} \sum_{n,p} \frac{1}{\xi_n^2 + \omega_p^2} = \sum_n \frac{g_{\mathrm{eff}} \kappa^2}{\beta} \frac{1}{(2\pi)^3} \int \mathrm{d}^3 p \frac{1}{\xi_n^2 + \omega_p^2} . \tag{3.2}$$

Introducing density of states $\sigma(\omega)$ from the ansatz

$$\frac{1}{(2\pi)^3} \int \mathrm{d}^3 p \equiv \int \mathrm{d}\omega_p \, \sigma(\omega_p) , \tag{3.3}$$

we get

$$A = \sum_n \frac{g_{\mathrm{eff}} \kappa^2}{\beta} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega \, \sigma(\omega)}{\xi_n^2 + \omega^2} \approx \sum_n \frac{g_{\mathrm{eff}} \kappa^2}{\beta} \sigma(0) \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\xi_n^2 + \omega^2} , \tag{3.4}$$

where $\omega_p = p^2/2m - \mu = 0$ is the Fermi surface and we have approximated by using density of states at the Fermi surface by putting a cut-off $n_{\max}$

$$\omega_D = k_B T \pi (2n_{\max} + 1) . \tag{3.5}$$

Thus

$$A = g_{\text{eff}}\kappa^2\sigma(0)\ln\frac{4\gamma\omega_D}{\pi k_B T}, \quad \gamma = \text{the Euler constant}. \tag{3.6}$$

The coefficient of $|\phi_c|^2$ terms $\Gamma[\phi_c, \phi_c^*]$ is then

$$\kappa^2\left(1 - g_{\text{eff}}\sigma(0)\ln\frac{4\gamma\omega_D}{\pi k_B T}\right)\phi_c^*\phi_c . \tag{3.7}$$

The expression within the brackets gives the critical temperature as

$$T_c = \frac{4\gamma\omega_D}{\pi k_B}\exp\left(-\frac{1}{g_{\text{eff}}\sigma(0)}\right). \tag{3.8}$$

Note that the terms appearing before the logarithm in (3.7) appear in the exponent while terms which are arguments for the log appear as $A_D\omega$ in the formula for T_c.

Since the Higgs mechanism is the relativistic version of the above, one may write the entire calculation of one-loop diagrams as well as the calculation of T_c in the form of corrections to the calculation of mass of the Higgs field. We shall adopt this procedure with slight variations.

4. The point we wish to make is summarized by (3.8) which has the form $A\omega_D \exp[-2/g_{\text{eff}}\sigma(0)]$. In order that the exponential factor does not present too much of a restriction $1/g_{\text{eff}}\sigma(0)$ should be of the order of unity. This gives $T_c \approx (\gamma\omega_D/\pi k_B)\exp(-2)$.

Before proceeding, we shall make some remarks about the amino acids. The crystalline structure of amino acids is characterized by the graph shown in fig. 1.

Apart from the N terms, here R_1 to R_4 are the residues which specify the amino acid concerned. The peptide bond which gives rise to proteins is formed, for example, where the molecules of water are consistently expelled and the lattice structure is reduced to the simple form

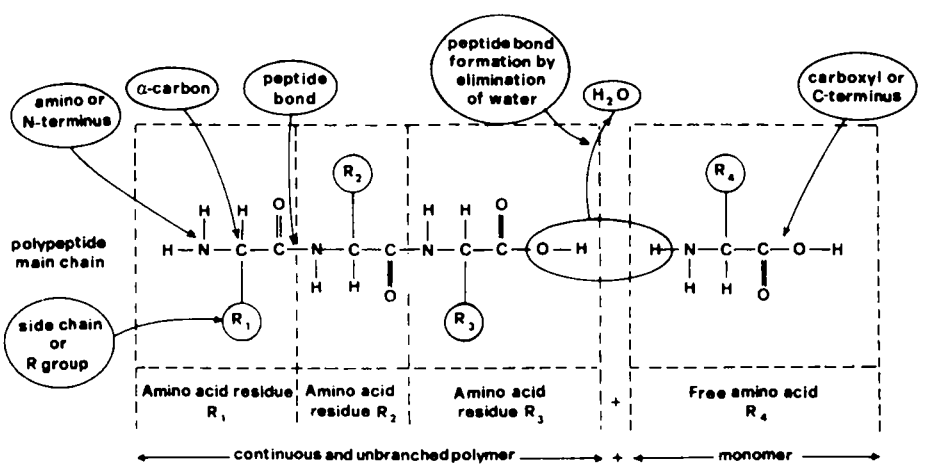
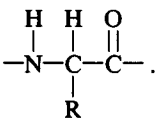


Fig. 1. This diagram is taken from ref. [4].



The hydrogens give up their loose electrons and act as metallic hydrogens. (Superconductivity has been established for a similar case by Carr [5].)

5. *Helicity and chirality of the matter fields* [#1]. Let us review the definition of chirality, which is the eigenvalue of γ_5, with $\gamma_5=+1$ corresponding to right-handedness, and $\gamma_5=-1$ to left-handedness:

$$\gamma_5 R=R, \quad \bar{R}\gamma_5=-\bar{R}, \qquad \gamma_5 L=-L, \quad \bar{L}\gamma_5=\bar{L}, \tag{5.1}$$

where R, L are Dirac spinors with only two independent components. They may be obtained from a four-component Dirac spinor ψ by the following projections:

$$R=\tfrac{1}{2}(1+\gamma_5)\psi, \quad \bar{R}=\tfrac{1}{2}\bar{\psi}(1-\gamma_5), \qquad L=\tfrac{1}{2}(1-\gamma_5)\psi, \quad \bar{L}=\tfrac{1}{2}\bar{\psi}(1+\gamma_5)\,. \tag{5.2}$$

For later applications, it is important to note that

$$\bar{\psi}\psi=\bar{L}R+\bar{R}L, \quad \bar{\psi}\gamma^\mu\psi=\bar{L}\gamma^\mu L+\bar{R}\gamma^\mu R\,, \tag{5.3}$$

written as

$$(\alpha\cdot p+\beta m)\psi=E\psi, \quad E=(p^2+m^2)^{1/2}\,. \tag{5.4}$$

Using the identity $\alpha=\gamma_5\sigma$, and the fact that γ_5 and σ commute, we can rewrite this in the form

$$\sigma\cdot\hat{p}R=\frac{E}{p}R-\frac{m}{p}\beta L, \quad p=|p|, \qquad \sigma\cdot\hat{p}L=-\frac{E}{p}L-\frac{m}{p}\beta R\,. \tag{5.5}$$

These equations become decoupled if $m=0$:

$$\sigma\cdot\hat{p}R=R, \quad \sigma\cdot\hat{p}L=-L\,. \tag{5.6}$$

Therefore, for massless Dirac particles, chirality is the same as helicity, for antiparticles, chirality is the opposite of helicity. [An antiparticle has the same chirality as the particle, by definition; but it has the opposite helicity due to a change in the sign of E in (5.4).]

A conventional mechanical mass term in the lagrangian density cannot be invariant under SU(2), because it is proportional to $\bar{\psi}\psi=\bar{L}R+\bar{R}L$. Therefore, in this theory the electron mass can arise only by virtue of a spontaneous breakdown of SU(2). A convenient way to implement this is to introduce a doublet Higgs field

$$\phi=\begin{pmatrix}\phi_+\\ \phi_0\end{pmatrix}, \tag{5.7}$$

where the subscripts refer to the electric charges. The mass term is

$$\mathcal{L}^{\text{mass}}=\vartheta\left(\bar{L}\phi R+\bar{R}\phi^\dagger L\frac{\rho}{\rho_0}\right), \tag{5.8}$$

where $\bar{L}\phi$ is an SU(2) singlet, and a Dirac spinor. If ϕ has non-zero vacuum value, then for low excitations and $\rho=\rho_0$, (5.8) is indistinguishable from a conventional mass term.

The Weinberg–Salam model is obtained by gauging SU(2)×U(1), generated by weak hypercharge.

[#1] This chapter is taken essentially from the book of Huang [6].

With these, the covariant derivative can be written in the form

$$D^{\mu}=\partial^{\mu}+ig(W_1^{\mu}t_1+W_2^{\mu}t_2)+ieQA^{\mu}+ieQ'Z_{\mu}\,. \tag{5.9}$$

Electric charge $=eQ$, $Q=\tau_3+\tau_0$ where the neutral charge matrix Q' is defined by

$$eQ'=\tau_3\cot\theta_w-\tau_0\tan\theta_w\,. \tag{5.10}$$

Here τ_0 is the singlet matrix among τ_0, τ_1, τ_2, τ_3, while $Q=\tau_3+\tau_0$.

To study the masses of the gauge fields, it is convenient to go to unitary gauge [#2], in which

$$\phi=\begin{pmatrix}0\\ \rho/\rho_0\end{pmatrix}, \tag{5.11}$$

where ρ/ρ_0 is a real field.

In terms of fields in the unitary gauge, the lagrangian density is

$$\mathcal{L}=-\tfrac{1}{4}(\boldsymbol{G}\cdot\boldsymbol{G}+\boldsymbol{H}\cdot\boldsymbol{H})+\tfrac{1}{4}g^2\rho^2\left(W_1^2+W_2^2+\frac{Z^2}{\cos^2\theta_w}\right)+\partial\rho\cdot\partial\rho-\lambda(\rho^2-\rho_0^2)^2\,. \tag{5.12}$$

Rewrite in the unitary gauge, where η is the real Higgs field in unitary gauge, in which $\phi(x)$ has the form (5.12) and

$$\rho(x)=\rho_0+\eta(x)\,. \tag{5.13}$$

6. To compute any further, we must guarantee the g_{eff} in the effective lagrangian is a positive number corresponding to attraction. Further, the proton left behind migrates to the nitrogen so that the Z^0 meson before being annihilated gives rise to an interaction between the quarks contained inside the proton and inside the nitrogen. (That it has to be nitrogen rather than the carbon (next door) is empirically guaranteed by the "right" configuration which is imparted to the corresponding molecule which contains sugars [#3] and no nitrogens [#4].)

To consider Z^0 containing part of the lagrangian, write it as

[#2] We should in fact have a triplet of fields

$$\begin{pmatrix}W^+\\ Z^0\\ W^-\end{pmatrix}\qquad W^+=W_1-\frac{IW_2}{\sqrt{2}}\,,$$

likewise $W^-=W_1+IW_2/\sqrt{2}$ belonging to a vector representation of the internal group SU(2), while the Higgs particles are a complex $\frac{1}{2}$ representation of the same group SU(2). We have neglected W^+ and W^- throughout this paper because their contribution is small at low momenta.

[#3] We may consider 2, 8, 20, 28, 50, 82 and 126 as "magic numbers" for nuclei, provided spin–orbit coupling is taken into account. (See ref. [7].) Nuclei containing 2, 8, 20, 28, 50, 82 or 126 neutrons or protons are particularly stable. The detailed evidence supporting this point of view is discussed in ref. [8] with the fact of 20 coming out naturally. Goeppert Mayer goes on to consider in detail the protons and neutrons and the spin–orbit couplings in terms of a potential energy which has a shape somewhat between that of a square well and a three-dimensional isotropic oscillator. (See table 1.)

[#4] The discussion from now on is not exclusively relevant to the rest of this paper. The problem for the usual quark model is to see if the Pauli principle holds for quark assignments. This means that in the final analysis the quark model which substitutes quarks for nucleons will have quark spheres of influence reduced depending on the number of colours. This implies that certain discrepancies in the nucleon model are removed when quarks are taken into account. These discrepancies, for example, concern the spin of ${}_{11}Na^{12}$. The magnetic moment of this nucleus would indicate $p_{3/2}$ rather than $d_{3/2}$ orbit. That these considerations have something new to tell us is an important point in its favour. (We would urge very strongly that parity violation among the proteins used by this mechanism should be further examined in order that the Z^0 and its decay are properly considered.)

Table 1

Oscillator number	Square well	Spectral term	Spin term	Number of states	Shells	Total number
0	1s	1s	$1s_{1/2}$	2	2	2
			$1p_{1/2}$	4		
1	1p	2p	$1p_{3/2}$	2	6	8
	1d	3d	$1d_{3/2}$	6		
2			$1d_{3/2}$	4	12	
	2s	2s	$2s_{1/2}$	2		20

$$L_{\text{int}}(Z^0)=\frac{eZ^0}{\sin\theta\cos\theta}[(T_{3L}-\sin^2\theta J_{\text{em L+R}})]\,. \tag{6.1}$$

Here J_{em} is the electromagnetic current, T_{3L} is the left-handed third component of the weak isospin consisting of the (anomaly-free) combination of the proton and neutron (p, n), (or the quarks inside the proton or the neutron). Z^0 here requires a mass because of the spontaneous symmetry breaking implied by the last equation.

$$L^{\text{mass}}=g_{\text{eff}}[T_L^3-(4\sin^2\theta J_{\text{em L+R}})]L_{\text{int}}\frac{\rho}{\rho_0}\,. \tag{6.2}$$

Here g_{eff} is proportional to $(e^2/16\sin^2\theta\cos^2\theta)\eta_{\mu\nu}/(k^2-m_Z^2)$ where $\eta_{\mu\nu}=1,-1,-1,-1$. This is positive provided $m_Z^2\gg k^2$ and we consider the three space-rotation invariant part of the expression above.

We shall take $(1-4\sin^2\theta)\approx\frac{1}{13}$ with the present *empirical value* of the parameter $\sin^2\theta\approx 0.231$. We shall not take this quantity to equal zero as has been done by authors of refs. [8,9] in the hope that the renormalization group will give the "exact" value as $\frac{1}{4}$ (personal communication). We do not believe this will ever happen.

For the mass term we take the ansatz, as the symmetry breaking which is given by eq. (2.4), which follows on from (2.5) and (2.7) where κ is a constant with dimension of mass. We shall introduce a source for the ϕ field in the form of eq. (2.8). In terms of the quantity ($\langle\varphi\rangle^0=250$ GeV) the electron mass M_e turns out to be a very tiny number $\approx 2\times 10^{-6}$. Such a number is large only for the top quark if its mass is in excess of 100 GeV. Thus the term which gives rise to this, looks like $f\varphi\bar{t}_L t_R$ where $f\geqslant\frac{2}{5}$. Some physicists (like Y. Nambu) take this as the defining property of the field φ, i.e. φ is considered as a $\bar{t}t$ composite.

According to Lee and Drell, a proposal is considered according to which the masses of the fermions in the standard model are determined by dynamical symmetry breaking rather than being introduced as arbitrary parameters in the lagrangian, they are determined self-consistently by the requirement that the proper self-energy vanish the fermion mass shell. They find that in the one-loop approximation it is possible to generate a heavy top quark mass dynamically while the other fermions remain massless [10]. We find that non-zero solutions for m_t do exist and they are always greater than 70 GeV for all values of cutoffs $\Lambda=5\,m_z$, $10\,m_z$, $20\,m_z$ and is the only Higgs mass explored in the problem.

One may thus be led to conclude that, while a heavy top quark mass can be generated dynamically, the observed masses of the light fermion cannot be generated in this approach.

Recently, Ruiz-Altaba, Gonzalez and Vargas [11] extended this to a two-loop calculation predicting $m_t=124$ GeV, $m_H=234$ GeV and the weak mixing angle $\sin^2\theta_w\simeq 0.24$.

Our physical motivation is to remove the fermion masses arbitrary parameters of the standard model and treat them as parameters determined by the dynamics in the sub-TeV region. Rather than cancel the divergences, they impose and interpret the standard model $\mathscr{L}$ as an effective lagrangian in order to determine the physical masses. Thus using Sakita's formulation, we get the result

$$T_c=\frac{\langle\varphi\rangle}{10^3}\exp[-2/g_{\text{eff}}\sigma(1-4\sin^2\theta)]\approx 2.5\times 10^2\,\text{K}\,. \tag{6.3}$$

The exponential factor gives $\exp(-26) \approx 10^{-10}$. Assuming that $\sigma(0) \approx m_Z^2$ we obtain $g_{\text{eff}}\sigma(0) \approx 1$.

7. Now we go on to the supersymmetry case. The symmetry resolves the dilemma of hierarchy. In a theory possessing fermion–boson symmetry because of the presence of both fermionic and bosonic radiative loops, the radiative shifts for scalar fields are not quadratically divergent. In the limit of broken supersymmetry, the shifts in v^2 only depend logarithmically on the cutoff Λ.

To guarantee that $g_{\text{eff}}\sigma(0) \approx 1$, we must take some mass in the theory which is large enough. The simplest mass is obtained by taking supersymmetry such that

$$\phi_0 \to \phi_0 + \alpha[\Delta M^2] \ln \frac{\Lambda}{\phi_0}, \tag{7.1}$$

where ΔM^2 is the mass splitting between supersymmetry fermions and is of order 1 TeV [12] which gives $\Delta M^2/92\, m_z^2 \approx 1$ and thus $g_{\text{eff}}\sigma(0)$ of the order of unity once again.

This paper started with the possibility of defining a science of life based on atomic physics. What we have shown is that if the model is Z_0 it is possible to have the quark instead of the nucleons. The mass of Z_0 is ≈ 100 GeV, while the mass of the quarks is ≈ 200 GeV so that if Z_0 loops are considered the $t\bar{t}$ quarks go through these loops. The g_{eff} which is proportional to $m_{\bar{Z}}{}^2$ then is of the right magnitude to be able to let these quarks make composites which are manifested through their passage through these loops.

References

[1] B. Sakita, Quantum theory of many-variable systems and fields (World Scientific, Singapore, 1985) pp. 42–51.
[2] V.L. Ginzberg and L.D. Landau, Sov. Phys. JETP 20 (1950) 1064.
[3] L.P. Gorkov, Sov. Phys. JETP Z. (1958) 505; JETP 9 (1959) 1364.
[4] A.R. Rees and M.J.E. Sternberg, From cells to atoms – An illustrated introduction to molecular biology (Blackwell, Oxford, 1984) p. 13.
[5] W.J. Carr Jr., Phys. Rev. B 33 (1986) 1585.
[6] K. Huang, Quarks, leptons and gauge fields (World Scientific, Singapore, 1982).
[7] M. Goeppert Mayer, Phys. Rev. 78 (1950) 16.
[8] G.E. Tranter and A.J. MacDermott, Croatica Chem. Acta 62 (2A) (1989) 165.
[9] S.F. Mason and G.E. Tranter, Molec. Phys. 53 (1984) 1091.
[10] I.-H. Lee and S.D. Drell, in: Fermion masses in the standard model, M.A.B. Bég Memorial Volume, eds. A. Ali and P. Hoodbhoy (World Scientific, Singapore, 1991) p. 13.
[11] M. Ruiz-Altaba, B. Gonzalez and M. Vargas, preprint CERN-TH 5558/89, UGVA-DPT 89/11-638 (unpublished); see also discussion in I. Jack and D.R.T. Jones, Phys. Lett. B 234 (1990) 321.
[12] U. Amaldi, Scientific American (1991), and CERN preprint.

PHYSICAL REVIEW B VOLUME 48, NUMBER 5 1 AUGUST 1993-I

Generalized spin systems and σ models

S. Randjbar-Daemi
International Centre for Theoretical Physics, Trieste, Italy

Abdus Salam
International Centre for Theoretical Physics, Trieste, Italy
and Department of Theoretical Physics, Imperial College, London, United Kingdom

J. Strathdee
International Centre for Theoretical Physics, Trieste, Italy
(Received 4 November 1992; revised manuscript received 16 March 1993)

A generalization of the SU(2) spin systems on a lattice and their continuum limit to an arbitrary compact group G is discussed. The continuum limits are, in general, nonrelativistic σ-model-type field theories targeted on a homogeneous space G/H, where H contains the maximal torus of G. In the ferromagnetic case the equations of motion derived from our continuum Lagrangian generalize the Landau-Lifshitz equations with quadratic dispersion relation for small wave vectors. In the antiferromagnetic case the dispersion law is always linear in the long-wavelength limit. The models become relativistic only when G/H is a symmetric space. Also discussed are a generalization of the Holstein-Primakoff representation of the SU(N) algebra, the topological term, and the existence of the instanton-type solutions in the continuum limit of the antiferromagnetic systems.

I. INTRODUCTION

Spin systems have been studied for many years and continue to provide new insights into the behavior of quantum mechanical systems with many degrees of freedom.[1] In this paper the notion of spin is generalized to include the representations of groups larger than the three-dimensional rotation group.[2]

The usual spin system is a collection of SU(2) spin operators associated with the sites of a lattice and coupled to their neighbors. The Hamiltonian is

$$H=\frac{1}{2}\sum_{m,n}J_{mn}S^{(m)}\cdot S^{(n)}+\cdots , \tag{1.1}$$

where lattice sites are labeled by sets of integers, $m,n,\ldots$, and the spin operators satisfy the SU(2) commutation rules,

$$[S_\alpha^{(m)},S_\beta^{(n)}]=i\delta_{mn}\epsilon_{\alpha\beta\gamma}S_\gamma^{(n)} .$$

In (1.1) only quadratic terms are indicated but higher-order terms could be added. Nonisotropic versions of (1.1), in which the SU(2) symmetry is broken, may also be considered.

If the lattice is one dimensional and the spin operators are represented by Pauli matrices, then the problem can be solved exactly.[3] Otherwise, the rigorous results are only partial, and it becomes necessary to use approximate or numerical methods.[1] Two regimes are particularly suitable for approximation: large spins and long wavelengths.[4] In the large spin or *correspondence theory limit,* the spin operators are represented by unit three-vectors,

$$S_\alpha^{(m)}\sim s\phi_\alpha^{(n)} ,\quad \phi^{(n)}\cdot\phi^{(n)}=1 ,\quad s\gg 1 ,$$

and the system becomes classical. In the long-wavelength limit the system approaches a continuum-field theory appropriate for the study of low-energy excitations. When both approximations are used in conjunction the system is described by a nonlinear σ model,[5]

$$S_\alpha^{(m)}\sim s\phi_\alpha(x) ,$$

$$\mathcal{L}\sim\frac{1}{2f}(\partial\phi_\alpha)^2+\cdots .$$

Corrections to the *correspondence theory limit* can be computed in the form of a loop expansion in which the system is represented by a so-called "quantum" nonlinear σ model. It is widely believed that the quantum nonlinear σ model provides an accurate description of the long-wavelength, low-energy properties of the spin system, even in the small spin regime.[6]

A primary goal in the investigation of spin systems is to find indications of phase transitions and critical behavior. Thus, at sufficiently low temperatures, it may be asked, is there long-range order? Does one find spontaneous magnetization in the ground state? In the classical regime this is certainly true, but, when quantum effects are taken into account, the ground state may well be disordered. For example, it is known that for the one-dimensional ($D=1$) system there is no long-range order in the ground state. On the other hand, for $D=3$ there can be long-range (antiferromagnetic) order. This order is destroyed at some finite critical temperature. For $D=2$ the ground state has been shown to exhibit long-range antiferromagnetic order for $s\geq 1$ but the situation for $s=\frac{1}{2}$ is not clear, at least in the isotropic models. Although there cannot be any long-range order for $D=2$ at finite temperature, it is interesting to consider

 ©1993 The American Physical Society

the behavior of the correlation length as the temperature goes to zero. Using renormalization-group methods it is possible to distinguish two regimes according to the coupling strength.[6] If the coupling exceeds a certain critical value then the correlation length remains finite for $T\to 0$, indicating a disordered ground state. If the coupling is less than the critical value then the correlation length diverges exponentially, indicating long-range order in the ground state.[7]

Most of the work in this field is concerned with the SU(2) spin models, since it is physically motivated. Our aim here is to extend some of these ideas to "spin" systems based on an arbitrary Lie group. Instead of associating SU(2) generators with the sites of a lattice we shall use the generators from any classical Lie algebra. We shall not attempt to generalize any of the rigorous theorems from the SU(2) literature. Rather, our purpose at this stage is to develop general formalism for treating the long-wavelength and correspondence-theory limits. Our view is that while such models may not have any immediate physical application (in the second paper of Ref. 2, Affleck has argued that the $N\to 0$ replica limit of the σ models targeted on $U(2N)/U(N)\times U(N)$ may give the critical exponents of the localization transition in the quantum Hall effect; however in Ref. 9 he has questioned the approach to the $N\to 0$ limit), they may eventually serve to cast some light on general features of the SU(2) models by placing them in a broader context. Indeed Read and Sachdev[8] (and independently Affleck[9]) have used the fermionic realization of the $U(N)$ generators to obtain models targeted on $U(N)/U(N-M)\times U(M)$. They have applied various well-known field-theoretical techniques, such as $1/N$ expansion, to examine their models. Our construction gives a more general class of models and, in principle, can be examined by similar methods. Our method, which uses bosonic variables, makes it clear why the models targeted on symmetric spaces such as $U(N)/U(N-M)\times U(M)$ are bound to have space-time Lorentz invariance.

From a mathematical point of view the generalized models are interesting in themselves. In going to groups of rank >1 the structure becomes much richer. For instance, one may contemplate new varieties of long-range order, beyond the familiar ferromagnetic and antiferromagnetic. The SU(3) analogue of the $s=\frac{1}{2}$, SU(2) system would employ the triplets 3 and 3^*. In the ground state one might expect to find an orderly arrangement of these states on the lattice. On the other hand, the large quantum number or correspondence limit will lead to σ models on one or other of the manifolds, $SU(3)/SU(2)\times U(1)$,$SU(3)/U(1)\times U(1)$, etc., depending on the ground state.

To generalize the well-known SU(2) spin-wave formalism it is necessary first of all to generalize the Holstein-Primakoff representation of spin matrices. The approach followed here starts from the coherent-state method used by Haldane,[10] which is easy to generalize. In this method, the finite-dimensional vector spaces on which the spin matrices act are provided with an over-complete basis labeled by coordinates on a coset space. Transition amplitudes in this basis are represented by path integrals whose meaning becomes unambiguous in the limit of large quantum numbers, and which can be used to extract a correspondence limit Lagrangian. In principle, it would also be possible to use this Lagrangian in the sub-correspondence regime, applying canonical quantization procedures to find corrections. However, the underlying path integral is subject to ordering ambiguities that make the passage to quantum theory less straightforward. This problem is solved, at least in the examples we have examined, by extrapolating directly from the coherent basis expectation values of the generators to operator expressions whose correctness is then verified by using the canonical commutation rules. The procedure will be illustrated in Sec. VI for the case of CP^N, where a realization of the Holstein-Primakoff type is obtained for the generators of $SU(N+1)$.

At the classical level, where ordering problems are suppressed, it is possible to give an explicit expression for the Lagrangian. The dynamical variables in terms of which it is expressed can be interpreted as coordinates on a coset space, G/H, where H includes the Cartan subgroup or maximal torus of G. If H is precisely the Cartan subgroup then G/H is a flag manifold. The detailed structure of the Lagrangian depends on the coherent state basis from which one starts. This basis is generated by applying finite transformations, L_ϕ, belonging to the group G, to some chosen reference state, $|\Lambda\rangle$,

$$|\phi\rangle = L_\phi|\Lambda\rangle\ ,\quad L_\phi\in G\ .$$

The stability group H is defined as the set of transformations, which leave the reference state invariant, up to a phase

$$h|\Lambda\rangle = |\Lambda\rangle e^{i\psi(h)}\ ,\quad h\in H\subset G\ .$$

The coordinates ϕ^μ, $\mu=1,2,\ldots,\dim G/H$, may be chosen in any convenient way to parametrize representatives, L_ϕ, from the cosets G/H. The Cartan components of the spin connection of G/H are defined by the 1 form

$$L_\phi^{-1}dL_\phi = -A^j(\phi)H_j+\cdots\ ,$$

where the operators H_j are generators of the Cartan algebra. These are among the generators of the stability group H and their eigenvalues serve to label the reference state,

$$H_j|\Lambda\rangle = |\Lambda\rangle\Lambda_j\ .$$

The coherent basis system is established at each site on the lattice and their direct products define an over-complete basis for the Hilbert space of the model. It will be shown in Sec. II that the correspondence limit Lagrangian takes the form

$$L=\frac{\hbar}{i}\sum_n A_\mu(\phi_n)\partial_t\phi_n^\mu - H(\phi)\ , \tag{1.2}$$

where

$$A_\mu(\phi_n) = A_\mu^j(\phi_n)\Lambda_{nj}$$

and

$$H(\phi)=\langle\phi|H|\phi\rangle$$
$$=\frac{1}{2}\sum_{m,n}J_{mn}Q_\alpha(\phi_m)Q_\alpha(\phi_n)+\cdots ,$$

where the functions $Q_\alpha(\phi)$ are defined as coherent state expectation values of the generators of G,

$$Q_\alpha(\phi_m)=\langle\phi|Q_\alpha^{(m)}|\phi\rangle .$$

Generalized spin waves are obtained by examining weak excitations of the system described by (1.2). These are defined, at the classical level, when the coherent state at each site on the lattice is close to its reference value. If the Lagrangian is translation invariant then the usual Fourier techniques can be used to extract the spectrum. Some examples of this will be given in Sec. III where the question of classical stability is considered.

Still at the classical level, it is straightforward to make the restriction to long-wavelength configurations, turning the lattice model into a continuum field theory. What emerges is a kind of nonlinear σ model with fields targeted on the coset space G/H. The Lagrangian will be first order in the time derivative but second (and higher) order in the space derivatives. However, if the (classical) ground state is "antiferromagnetic" in the sense that the reference state weights Λ_{nj} alternate in sign across the lattice, then some of the dynamical variables are algebraic and can be eliminated to give a Lagrangian, which is second order in the time derivative. This Lagrangian is generally nonrelativistic, even when higher-order space derivatives are neglected, unless G/H happens to be a symmetric space. This will be discussed in Sec. IV.

In passing from the lattice to the continuum description, a term appears in the Lagrangian density (for the antiferromagnetic case), which is a total derivative. Such a term has no relevance at the classical level, since it makes no contribution to the equations of motion. It is only an artefact of the method and may be safely discarded. In the quantized continuum theory, on the other hand, such a term may not be negligible. If it makes a finite contribution to the action functional then it will affect the sum over configurations. Since our discussion, in Sec. IV, of the passage to the continuum theory is couched in classical terms—the factor ordering question is ignored—we cannot determine whether such terms are actually needed for the long-wavelength description of spin systems at the quantum level. Haldane observed this term in his treatment of SU(2) spin systems in one-space dimension ($D=1$), and he noted that it has a topological interpretation, giving an alternating sign factor in the sum over configurations when the spin is a half integer.[11,2] Within the limits of our discussion we have confirmed that this effect persists in the generalized antiferromagnetic $D=1$ spin systems. For $D>1$ our total derivative term does not seem to have a topological significance, and it has been argued that, at least for smooth field configurations, it should probably be suppressed in the quantum theory.[12] However, nonsmooth configurations such as the point singular "hedghogs" may induce interesting physical phenomena. In fact, following the work of Haldane in SU(2) case, Read and Sachdev[13] have shown that close to a phase boundary the SU(N) hedghogs, crucially influence the nature of transition from an ordered Néel state to a disordered phase. Topological aspects of the $D=1$ systems are discussed in Sec. V where first-order equations for generalized instantons are derived.

The classical nonlinear σ models obtained by this approach can be quantized in the usual way. It is not at all obvious that such quantum nonlinear σ models have much relevance to the original spin problem but, as mentioned above, there is a common belief in the SU(2) case that they are good for describing the long-wavelength, low-energy features of the spin problem. Some support for this view may perhaps derive from the standard treatment of quantized σ models by dimensional regularization. In such treatments it is prescribed that quantum corrections associated with factor ordering should be discarded. Recall that it is precisely these ordering contributions—defined by the Holstein-Primakoff operator valued expressions for the generators $Q_\alpha(\phi_n)$—which distinguish the subcorrespondence regime on the lattice. If they are indeed not relevant in the continuum limit then the quantum nonlinear σ model should be appropriate for the long-wave behavior of the lattice model.

Finally, it may be remarked that the loop expansions of the quantized theory comprise quantum corrections to the classical spin wave amplitudes. Since both $A_\mu(\phi)$ and $Q_\alpha(\phi)$ in (1.2) are linear in the weights Λ_j, it is clear that the loop expansions can be read as expansions in powers of $1/\Lambda$ if the couplings J_{mn} are redefined to absorb one power of Λ, $J_{mn}=J'_{mn}/\Lambda$. With this interpretation, the classical theory emerges when $\Lambda\to\infty$. This is what is meant by the expression, large quantum number of correspondence limit.

II. THE SEMICLASSICAL REGIME

Our purpose is to discuss the large quantum number, or correspondence limit of generalized spin systems. Typically, the Hamiltonian is given as a sum of terms in which operators associated with sites on a lattice are coupled,

$$H=\frac{1}{2}\sum_{m,n}J_{mn}^{\alpha\beta}Q_\alpha^{(m)}Q_\beta^{(n)}+\cdots , \tag{2.1}$$

where $m,n,\ldots$ label the sites and the coefficients $J_{mn}^{\alpha\beta}$ are coupling parameters. The matrices $Q_\alpha^{(m)}$ are generators of some representation of a Lie group G. They satisfy the commutation rules

$$[Q_\alpha^{(m)},Q_\beta^{(n)}]=\delta_{mn}c_{\alpha\beta}{}^\gamma Q_\gamma^{(n)} . \tag{2.2}$$

The couplings are assumed to be G invariant, i.e., proportional to the Killing metric,

$$J_{mn}^{\alpha\beta}=J_{mn}g^{\alpha\beta} , \tag{2.3}$$

and also translation invariant,

$$J_{mn}=J_{m-n} . \tag{2.4}$$

Higher order and/or noninvariant coupling terms could

be included in the Hamiltonian (2.1), but we shall not consider such possibilities.

We are not looking for exact solutions. Following Affleck,[2] our aim is to find a Lagrangian description, which can be used for semiclassical approximations, generalized spin waves, and which leads to various σ-model-type theories in the continuum limit. Our approach is to generalize the method used by Haldane.[10] As outlined in Sec. I this involves the introduction of an over-complete basis of coherent states in the finite-dimensional vector space on which the Q_α act, followed by the construction of a path integral representation for transition amplitudes in the coherent basis. In the SU(2) case the coherent states in question are associated with the points of the manifold SU(2)/U(1), and they describe the orientation of the spin vector in the correspondence theory limit. This picture generalizes easily to the cosets G/H.

Let G be a Lie group and H one of its subgroups. We are interested in those subgroups, which include at least the Cartan subgroup (maximal Abelian subgroup). If H is precisely the Cartan subgroup then G/H is a flag manifold, but we may consider larger, non-Abelian subgroups so that G/H becomes a subspace of the flag manifold.

Suppose that G is decomposed into left cosets with respect to the subgroup H. From each coset one can choose a representative element $L_\phi \in G$ where $\phi=\{\phi^\mu\}$ labels a point on the manifold G/H. This means that, for arbitrary $g \in G$, there is a map

$$\phi \to \phi' = \phi'(\phi, g)$$

defined by

$$gL_\phi = L_{\phi'} h \ , \tag{2.5}$$

where $h = h(\phi, g) \in H$. The detailed form of the functions $\phi'(\phi,g)$ and $h(\phi,g)$ depends on the choice of representative elements L_ϕ.

In the vector space that carries one of the irreducible representations of G, choose one vector $|\Lambda\rangle$ that is invariant, up to a phase, under the action of H,

$$h|\Lambda\rangle = |\Lambda\rangle e^{i\psi(h)} \ . \tag{2.6}$$

This vector will be referred to as the "reference state." Now, corresponding to the points $\phi \in G/H$, define the coherent states

$$|\phi\rangle = L_\phi |\Lambda\rangle \ . \tag{2.7}$$

Under the action of G these states are transformed according to

$$g|\phi\rangle = |\phi'\rangle e^{i\psi(h)} \ , \tag{2.8}$$

where ϕ' and h are determined by (2.5).

The coherent states can be expanded in an orthonormal basis,

$$|\phi\rangle = \sum_\lambda |\lambda\rangle\langle\lambda|L_\phi|\Lambda\rangle \ , \tag{2.9}$$

and it is possible to project the orthonormal basis vectors from the coherent states by integrating over the coset manifold,

$$|\lambda\rangle = \int d\mu(\phi) |\phi\rangle\langle\Lambda|L_\phi^{-1}|\lambda\rangle \ , \tag{2.10}$$

where $d\mu$ is a suitably normalized G invariant measure on G/H. The coherent states therefore constitute an over-complete basis.

Of particular interest is the overlap between neighboring states,

$$\begin{aligned}\langle\phi + d\phi|\phi\rangle &= \langle\Lambda|L_{\phi+d\phi}^{-1} L_\phi|\Lambda\rangle \\ &= \langle\Lambda|(1 - L_\phi^{-1} dL_\phi)|\Lambda\rangle \\ &= 1 + A^j \Lambda_j \ ,\end{aligned} \tag{2.11}$$

where the 1 forms, $A^j = d\phi^\mu A_\mu^j(\phi)$, are the Cartan components of the spin connection on G/H. To obtain this we have expanded the Maurer-Cartan form in a basis of the Lie algebra of G,

$$\begin{aligned}L_\phi^{-1} dL_\phi &= e^\alpha Q_\alpha \\ &= -A^j H_j + e^{\bar\alpha} Q_{\bar\alpha} + e^{\dot\alpha} Q_{\dot\alpha} \ ,\end{aligned} \tag{2.12}$$

where the operators Q_α satisfy the commutation rules

$$[Q_\alpha, Q_\beta] = c_{\alpha\beta}{}^\gamma Q_\gamma \ . \tag{2.13}$$

The generators of the Cartan algebra are denoted H_j. These operators are all in the algebra of H together (in general) with non-Abelian elements $Q_{\bar\alpha}$. The remaining generators are denoted by $Q_{\dot\alpha}$. Their coefficient 1 forms, $e^{\dot\alpha}$, in the Maurer-Cartan form define the frames on G/H. The requirement (2.6) that $|\Lambda\rangle$ be invariant, up to a phase, under the action of H means

$$\begin{aligned}&H_j|\Lambda\rangle = |\Lambda\rangle\Lambda_j \ , \\ &Q_{\bar\alpha}|\Lambda\rangle = 0 \ .\end{aligned} \tag{2.14}$$

The operators $Q_{\dot\alpha}$ belong to some representation of H. This representation is often reducible, but it cannot contain any components that are neutral with respect to the Cartan algebra. Hence, $\langle\Lambda|Q_{\dot\alpha}|\Lambda\rangle = 0$ and the result (2.11) follows.

The infinitesimal formula (2.11) can be integrated to obtain a path integral representation for the finite overlap. To do this one makes repeated use of the completeness condition (2.10) or

$$1 = \int d\mu(\phi) |\phi\rangle\langle\phi| \ . \tag{2.15}$$

Thus, one writes firstly

$$\langle\phi'|\phi\rangle = \int \langle\phi_N|\phi_{N-1}\rangle d\mu(\phi_{N-1}) \langle\phi_{N-1}|\phi_{N-2}\rangle d\mu(\phi_{N-2}) \cdots d\mu(\phi_1)\langle\phi_1|\phi_0\rangle \ ,$$

where $\phi_N=\phi'$ and $\phi_0=\phi$. The intermediate points, $\phi_{N-1},\phi_{N-2},\ldots,\phi_1$ are integrated over G/H. In the limit $N\to\infty$ this gives rise in the usual formal way to a path integral over configurations $\phi(t)$. The factors in the integrand are given by (2.11),

$$\langle\phi+d\phi|\phi\rangle=\exp(A^j\Lambda_j)$$
$$=\exp(dt\,\dot\phi^\mu A_\mu^j(\phi)\Lambda_j)\,,$$

to first order in dt. Hence, the finite overlap is

$$\langle\phi'|\phi\rangle=\int(d\mu)\exp\int dt\,\dot\phi^\mu A_\mu^j(\phi)\Lambda_j\,. \tag{2.16}$$

Finally, the coherent state transition amplitudes are represented by the path integrals

$$\left\langle\phi'\left|\exp\left[-\frac{i}{\hbar}tH\right]\right|\phi\right\rangle=\int(d\mu)\exp\left[\frac{i}{\hbar}\int dt\,L\right]\,, \tag{2.17}$$

where the Lagrangian is given by

$$L=\frac{\hbar}{i}\dot\phi^\mu A_\mu^j(\phi)\Lambda_j-H(\phi)\,. \tag{2.18}$$

The classical Hamiltonian $H(\phi)$ is defined by the diagonal elements of H in the coherent basis,

$$H(\phi)=\langle\phi|H|\phi\rangle\,. \tag{2.19}$$

The Euler Lagrange equations deriving from (2.18) take the form

$$\frac{\hbar}{i}\Lambda_jF_{\mu\nu}^j\dot\phi^\nu=\frac{\partial H}{\partial\phi^\mu}\,, \tag{2.20}$$

where $F_{\mu\nu}^j=\partial_\mu A_\nu^j-\partial_\nu A_\mu^j$ are the Cartan components of the curvature tensor on G/H.

The path integral representation (2.17) is of course only formal. To give it a precise meaning one must apply canonical quantization methods to the classical Lagrangian (2.18) and invent a suitable prescription for interpreting the ambiguous factor ordering in $H(\phi)$. The canonical momenta are defined by

$$\pi_\mu=\frac{\partial L}{\partial\dot\phi^\mu}=\frac{\hbar}{i}A_\mu^j(\phi)\Lambda_j\,. \tag{2.21}$$

To proceed further with this it will be necessary to choose some parametrization of G/H and obtain explicit formulae for the components A_μ^j. Once the commutation rules have been determined for the ϕ^μ, it becomes possible to consider the structure of $H(\phi)$. For the systems we are concerned with, the Hamiltonian is given as a polynomial in the generators Q_α. This means that the classical Hamiltonian will be expressed in terms of the functions

$$Q_\alpha(\phi)=\langle\phi|Q_\alpha|\phi\rangle\,. \tag{2.22}$$

These functions are unambiguous at the classical level but when the dynamical variables ϕ^μ are quantized it is necessary to determine the factor ordering such that the algebra is realized,

$$[Q_\alpha(\phi),Q_\beta(\phi)]=c_{\alpha\beta}{}^\gamma Q_\gamma(\phi)\,.$$

This is precisely the problem solved by Holstein and Primakoff[14] for the case of SU(2). In Sec. VI we shall give a generalization for the case of SU($N+1$) with coordinates ϕ^μ on CP^N. There it will be verified that the ϕ commutators vanish like $1/\Lambda$ in the limit $\Lambda\to\infty$ [see Eq. (6.10)]. But first we shall continue with the classical approximation (i.e., the large quantum number limit).

III. WEAK EXCITATIONS AND CLASSICAL STABILITY

Suppose that the Hamiltonian for the generalized spin system on a lattice is given as a sum over pairs of sites, etc., as described in Sec. I. In the classical approximation,

$$H(\phi)=\frac{1}{2}\Sigma J_{mn}g^{\alpha\beta}Q_\alpha(\phi_m)Q_\beta(\phi_n)+\cdots\,, \tag{3.1}$$

where $g^{\alpha\beta}$ is the Killing metric and the functions $Q_\alpha(\phi)$ are defined by (2.22). It is convenient to express these functions in terms of matrix elements of L_ϕ in the adjoint representation. These matrices, $D_\alpha{}^\beta$, are defined by

$$g^{-1}Q_\alpha g=D_\alpha{}^\beta(g)Q_\beta\,,\quad g\in G\,.$$

It follows from the definition (2.7) of the coherent states that

$$Q_\alpha(\phi)=\langle\Lambda|L_\phi^{-1}Q_\alpha L_\phi|\Lambda\rangle$$
$$=D_\alpha{}^\beta(L_\phi)\langle\Lambda|Q_\beta|\Lambda\rangle$$
$$=D_\alpha{}^j(L_\phi)\Lambda_j\,. \tag{3.2}$$

Hence (3.1) takes the form

$$H(\phi)=\tfrac{1}{2}\Sigma J_{mn}g^{\alpha\beta}D_\alpha{}^j(L_{\phi_m})D_\beta{}^k(L_{\phi_n})\Lambda_{mj}\Lambda_{nk}+\cdots$$
$$=\tfrac{1}{2}\Sigma J_{mn}\Lambda_m^jD_j{}^k(L_{\phi_m}^{-1}L_{\phi_n})\Lambda_{nk}+\cdots\,, \tag{3.3}$$

where the invariance of the metric has been used,

$$g^{\beta\alpha}D_\alpha{}^j(L)=g^{j\alpha}D_\alpha{}^\beta(L^{-1})=g^{jk}D_k{}^\beta(L^{-1})\,.$$

The indices j,k refer to components in the Cartan algebra, and we take a basis such that the metric tensor is block diagonal.

Since the coset manifold is homogeneous, no generality is lost by assuming that ϕ_n becomes independent if n in the ground state, i.e.,

$$\langle L_{\phi_m}^{-1}L_{\phi_n}\rangle=1\,. \tag{3.4}$$

With this convention the ground state value of the Hamiltonian (3.3) reduces to

$$\langle H(\phi)\rangle=\tfrac{1}{2}\sum_{mn}J_{mn}\Lambda_m\cdot\Lambda_n+\cdots\,. \tag{3.5}$$

Minimization of this expression must be our guide in choosing the weights Λ_n that define the ground-state configuration. To be more specific, we shall suppose that the pattern of weights takes the form of a finite number of translation invariant sublattices. Write the variables in the form

$$\{\phi_n\}=\{\phi_{\underline{n}\nu}\}\ , \tag{3.6}$$

where $\underline{n}=(n_1,n_2,\ldots,n_D)$ is a D vector with integer components and $\nu=1,\ldots,f$ indicates the sublattice. The weights are independent of $\underline{n}$,

$$\Lambda_{\underline{n}\nu}=\Lambda_\nu\ .$$

The translation invariant coupling parameters $J_{\underline{m}\mu,\underline{n}\nu}$ depend on $\underline{m}-\underline{n}$. The ground-state energy per cell is then given by

$$\tfrac{1}{2}\sum_{\underline{n},\nu,\nu} J_{\underline{n}\nu,\underline{0}\nu}\Lambda_\nu\cdot\Lambda_\nu\ . \tag{3.7}$$

To minimize the ground-state energy density in, for example, a one-dimensional system with nearest-neighbor couplings one should choose $\Lambda_n\cdot\Lambda_{n+1}$ to be positive (negative) if $J_{n,n+1}$ is negative (positive). In this way one is led to the most obvious generalizations of ferromagnetic (antiferromagnetic) order. In more than one dimension there are new possibilities. For example, in two dimensions with a triangular lattice each site has six nearest neighbors. If all nearest neighbors are coupled with the same positive strength then one can make an ordered ground state with three distinct sublattices such that the energy density is

$$J(\Lambda_1\Lambda_2+\Lambda_2\Lambda_3+\Lambda_3\Lambda_1)$$
$$=\frac{J}{2}[(\Lambda_1+\Lambda_2+\Lambda_3)^2-\Lambda_1^2-\Lambda_2^2-\Lambda_3^2]\ .$$

Hence the configuration with $\Lambda_1+\Lambda_2+\Lambda_3=0$ would be favored. Such a generalized antiferromagnet becomes a possibility in the SU(3) models.

To test the stability of conjectured ground-state configurations one can compute the energy associated with weak perturbations and find the excitation spectrum. Write

$$\phi_n=\phi+\Delta\phi_n\ , \tag{3.8}$$

where ϕ is a constant and $\Delta\phi_n$ is small. Substitute into (3.3) and collect the bilinear terms. To carry out this computation we need the formula

$$L_\phi^{-1}L_{\phi+\Delta\phi}=1+\Delta\phi^\mu L^{-1}\partial_\mu L+\tfrac{1}{2}\Delta\phi^\mu\Delta\phi^\nu L^{-1}\partial_\mu\partial_\nu L+\cdots$$
$$=1+\Delta\phi^\mu e_\mu{}^\alpha Q_\alpha+\tfrac{1}{2}\Delta\phi^\mu\Delta\phi_\nu(\partial_\mu e_\nu{}^\alpha Q_\alpha+e_\mu{}^\alpha e_\nu{}^\beta Q_\alpha Q_\beta)+\cdots\ ,$$

which derives from (2.12). Hence,

$$L_{\phi_m}^{-1}L_{\phi_n}=1+[(\Delta\phi_n{}^\mu-\Delta\phi_m{}^\mu)e_\mu{}^\alpha+\tfrac{1}{2}(\Delta\phi_n{}^\mu\Delta\phi_n{}^\nu-\Delta\phi_m{}^\mu\Delta\phi_m{}^\nu)\partial_\mu e_\nu{}^\alpha-\tfrac{1}{2}\Delta\phi_m{}^\mu\Delta\phi_n{}^\nu e_\mu{}^\beta e_\nu{}^\gamma c_{\beta\gamma}{}^\alpha]Q_\alpha$$
$$+\tfrac{1}{2}(\Delta\phi_n{}^\mu-\Delta\phi_m{}^\mu)(\Delta\phi_n{}^\nu-\Delta\phi_m{}^\nu)e_\mu{}^\alpha e_\nu{}^\beta Q_\alpha Q_\beta+\cdots\ ,$$

and, in the adjoint representation,

$$D_j{}^k(L_{\phi_m}^{-1}L_{\phi_n})=\delta_j^k+\tfrac{1}{2}(\Delta\phi_n{}^\mu-\Delta\phi_m{}^\mu)(\Delta\phi_n{}^\nu-\Delta\phi_m{}^\nu)e_\mu{}^\alpha e_\nu{}^\beta c_{j\alpha}{}^\gamma c_{\gamma\beta}{}^k+\cdots\ .$$

With this approximation the Hamiltonian (3.3) takes the form

$$H(\phi)=\tfrac{1}{2}\Sigma J_{mn}[\Lambda_m\cdot\Lambda_n+\tfrac{1}{2}(\Delta\phi_n{}^\mu-\Delta\phi_m{}^\mu)(\Delta\phi_n{}^\nu-\Delta\phi_m{}^\nu)k_{\mu\nu}^{mn}(\phi)+\cdots]\ , \tag{3.9}$$

where $k_{\mu\nu}$ is a symmetric G-invariant tensor on G/H defined by

$$k_{\mu\nu}^{mn}=e_\mu{}^\alpha e_\nu{}^\beta\Lambda_m^j c_{j\alpha}{}^\gamma c_{\gamma\beta}{}^k\Lambda_{nk}\ . \tag{3.10}$$

The ground state will be stable against weak classical perturbations if the matrix, $J_{mn}k_{\mu\nu}^{mn}$, is positive. To examine this question it is helpful to express the algebra of G in the Cartan-Weyl basis with generators H_j and E_α, where α denotes a root. The commutation rules include

$$[H_j,E_{\pm\alpha}]=\pm\alpha_j E_{\pm\alpha}\ ,$$
$$[E_\alpha,E_{-\alpha}]=\alpha^j H_j\ .$$

Since we are dealing with unitary representations we can choose the basis such that

$$H_j{}^\dagger=H_j\quad\text{and}\quad E_\alpha{}^\dagger=E_{-\alpha}\ .$$

The root vectors are then real. However, it is important to remark that, since $L^{-1}dL$ is anti-Hermitian, the frame components in the Cartan-Weyl basis must satisfy

$$(\Delta\phi^\mu e_\mu{}^\alpha)^*=-\Delta\phi^\mu e_\mu{}^{-\alpha}\ .$$

The tensor (3.10) reduces in this basis to a sum over roots,

$$k_{\mu\nu}^{mn}=\sum_{\text{roots}} e_\mu{}^\alpha e_\nu{}^{-\alpha}\Lambda_m\cdot\alpha\alpha\cdot\Lambda_n\ . \tag{3.11}$$

Therefore,

$$\Delta\phi^\mu\Delta\phi^\nu k_{\mu\nu}^{mn}=-\sum_{\text{roots}}|\Delta\phi^\mu e_\mu{}^\alpha|^2\Lambda_m\cdot\alpha\alpha\cdot\Lambda_n\ .$$

It follows that the conjectured ground state will be stable if the matrix, $J_{mn}\Lambda_{mj}\Lambda_{nk}$, is negative definite. This is a sufficient condition. It would be straightforward, in principle, to sharpen this criterion for a specific model, but it would not be very useful to pursue the question in generality.

We conclude this discussion with brief remarks about the excitation spectrum. To simplify the notation define the frame components of the weak fluctuations,

$$\Delta\phi_n{}^\mu e_\mu{}^\alpha=\psi_n^\alpha=-(\psi_n^{-\alpha})^*\ . \tag{3.12}$$

In terms of these variables the bilinear part of the Hamiltonian (2.31) becomes

$$H_2=-\tfrac{1}{2}\sum_{\alpha>0}\sum_{m,n}J_{mn}\Lambda_m\cdot\alpha\alpha\cdot\Lambda_n|\psi_n^\alpha-\psi_m^\alpha|^2\,, \tag{3.13}$$

where the sum is restricted to positive roots. With the translation invariant background, described above, comprising a finite number of sublattices, $\{n\}=\{\underline{n},\nu\}$, it is natural to use Fourier series. Define the Fourier components,

$$\begin{aligned}\tilde{\psi}_\nu^\alpha(\underline{k})&=\sum_{\underline{n}}\psi_{\underline{n}\nu}^\alpha e^{-i\underline{k}\cdot\underline{n}}\,,\\ \psi_{\underline{n}\nu}^\alpha&=\int\left[\frac{dk}{2\pi}\right]^D\tilde{\psi}_\nu^\alpha(\underline{k})e^{i\underline{k}\cdot\underline{n}}\,,\end{aligned} \tag{3.14}$$

where the integration ranges over a cell of volume $(2\pi)^D$. The Hamiltonian (3.13) then takes the form

$$\begin{aligned}H_2&=-\frac{1}{2}\sum_{\alpha>0}\int\left[\frac{dk}{2\pi}\right]^D\sum_{\underline{n}}\sum_{\nu,\nu'}J_{\underline{n}\nu,\underline{0}\nu'}\Lambda_\nu\cdot\alpha\alpha\cdot\Lambda_{\nu'}|\tilde{\psi}_\nu^\alpha(\underline{k})e^{i\underline{k}\cdot\underline{n}/2}-\tilde{\psi}_{\nu'}^\alpha(\underline{k})e^{-i\underline{k}\cdot\underline{n}/2}|^2\\ &=\sum_{\alpha>0}\int\left[\frac{dk}{2\pi}\right]^D\sum_{\nu,\nu'}\tilde{\psi}_\nu^\alpha(\underline{k})^*H_{\nu\nu'}^\alpha(\underline{k})\tilde{\psi}_{\nu'}^\alpha(\underline{k})\,.\end{aligned} \tag{3.15}$$

Stability requires that the matrix $H_{\nu\nu'}^\alpha(\underline{k})$ should be positive definite for each root.

To obtain the frequency spectrum it is necessary to extract the bilinear part of the Lagrangian (2.18),

$$L=\frac{\hbar}{i}\sum_n\dot{\phi}_n^\mu A_\mu^j(\phi_n)\Lambda_{nj}-H(\phi)\,.$$

Substituting (3.8) and discarding a total derivative gives

$$\begin{aligned}L_2&=\frac{\hbar}{2i}\sum_n\Delta\dot{\phi}_n^\mu\Delta\phi_n^\nu F_{\nu\mu}^j(\phi)\Lambda_{nj}-H_2\\ &=\sum_{\alpha>0}\left[\frac{\hbar}{2i}\sum_n\alpha\cdot\Lambda_n\psi_n^{\alpha*}(\overleftarrow{\partial}_t-\overrightarrow{\partial}_t)\psi_n^\alpha\right.\\ &\qquad\left.+\frac{1}{2}\sum_{mn}J_{mn}\Lambda_m\cdot\alpha\alpha\cdot\Lambda_n|\psi_n^\alpha-\psi_m^\alpha|^2\right]\,.\end{aligned} \tag{3.16}$$

To obtain this expression we have used the Maurer-Cartan expression for the curvature tensor,

$$F_{\mu\nu}{}^j=e_\mu{}^\alpha e_\nu{}^\beta c_{\alpha\beta}{}^j\,.$$

In the Cartan-Weyl basis this gives

$$F_{\mu\nu}{}^j\Lambda_{nj}=\sum_{\text{roots}}e_\mu{}^\alpha e_\nu{}^{-\alpha}(\alpha\cdot\Lambda_n)\,. \tag{3.17}$$

The linearized equations of motion are easily obtained,

$$-\alpha\cdot\Lambda_\nu\frac{\hbar}{i}\partial_t\psi_\nu^\alpha(\underline{k})=\sum_{\nu'}H_{\nu\nu'}^\alpha(\underline{k})\psi_{\nu'}^\alpha(\underline{k})\,.$$

The frequencies are therefore given by the zeroes of the determinant

$$\det[\alpha\cdot\Lambda_\nu\delta_{\nu\nu'}\hbar\omega-H_{\nu\nu'}^\alpha(\underline{k})]\,. \tag{3.18}$$

To illustrate, we consider the one-dimensional chain with nearest-neighbor coupling where there are two familiar cases, ferromagnetic and antiferromagnetic, for which the spectrum is easily obtained.

A. Ferromagnetic ground state ($J<0$)

In this case we have a simple lattice with Λ independent of n. The fluctuation Hamiltonian (3.13) takes the form

$$\begin{aligned}H_2&=-J\sum_{\alpha>0}(\alpha\cdot\Lambda)^2\sum_n|\psi_n^\alpha-\psi_{n+1}^\alpha|^2\\ &=-J\sum_{\alpha>0}(\alpha\cdot\Lambda)^2\int\frac{dk}{2\pi}\tilde{\psi}^\alpha(k)^*|e^{ik/2}-e^{-k/2}|^2\psi^\alpha(k)\,,\end{aligned}$$

i.e.,

$$H^\alpha(k)=-4J(\alpha\cdot\Lambda)^2\sin^2k/2\,, \tag{3.19}$$

so that

$$\hbar\omega=-4J\alpha\cdot\Lambda\sin^2k/2\,. \tag{3.20}$$

The sign of ω is not significant. If $\alpha\cdot\Lambda<0$ we can interpret ψ^α as the creation operator and $\psi^{-\alpha}$ as the annihilation operator for an excitation of energy $\hbar|\omega|$.

B. Antiferromagnetic ground state ($J>0$)

In this case we have two sublattices with $\Lambda_1=\Lambda=-\Lambda_2$. The Hamiltonian (3.13) now takes the form

$$\begin{aligned}H_2&=J\sum_{\alpha>0}(\alpha\cdot\Lambda)^2\sum_n(|\psi_{n1}^\alpha-\psi_{n2}^\alpha|^2+|\psi_{n1}^\alpha-\psi_{n-1,2}^\alpha|^2)\\ &=J\sum_{\alpha>0}(\alpha\cdot\Lambda)^2\int\frac{dk}{2\pi}[|\psi_1^\alpha(k)-\psi_2^\alpha(k)|^2\\ &\qquad+|e^{ik/2}\psi_1^\alpha(k)-e^{-ik/2}\psi_2^\alpha(k)|^2]\,,\end{aligned}$$

i.e.,

$$H^\alpha(k)=2J(\alpha\cdot\Lambda)^2\begin{bmatrix}1 & -e^{-ik/2}\cos k/2\\ -e^{ik/2}\cos k/2 & 1\end{bmatrix}\,. \tag{3.21}$$

The secular determinant (3.18) is

$$\det\begin{vmatrix} \alpha\cdot\Lambda\hbar\omega-2J(\alpha\cdot\Lambda)^2 & 2J(\alpha\cdot\Lambda)^2e^{-ik/2}\cos k/2 \\ 2J(\alpha\cdot\Lambda)^2e^{ik/2}\cos k/2 & -\alpha\cdot\Lambda\hbar\omega-2J(\alpha\Lambda)^2 \end{vmatrix}$$
$$=(\alpha\cdot\Lambda)^2[-\hbar^2\omega^2+4J^2(\alpha\cdot\Lambda)^2\sin^2 k/2]$$

so that

$$\hbar\omega=\pm 2J(\alpha\cdot\Lambda)\sin k/2\ . \tag{3.22}$$

In contrast to the ferromagnetic case (3.20), the spectrum is linear in k near $k=0$. For symmetric spaces like $SU(N+M)/SU(N)\times SU(M)\times U(1)$, it can be shown that $\alpha\cdot\Lambda$ is independent of α and hence, so are the frequencies. The spectrum is relativistic for small k. In this case the corresponding low-energy-field theory will turn out to be the well-known relativistic nonlinear σ model targeted on the appropriate symmetric space.

IV. THE CONTINUUM LIMIT

To extract a continuum description of the long-wavelength behavior of the system one supposes that the dynamical variables are slowly varying across the lattice

$$\phi_m-\phi_n\ll\phi_n\ .$$

In the correspondence theory limit this means

$$Q_\alpha(\phi_m)Q^\alpha(\phi_n)\simeq\Lambda_m^j D_j{}^k(L_{\phi_m}^{-1}L_{\phi_n})\Lambda_{nk}$$
$$=\Lambda_m\cdot\Lambda_n+\tfrac{1}{2}(\phi_m-\phi_n)^\mu(\phi_m-\phi_n)^\nu k_{\mu\nu}^{mn}(\phi_n)+\cdots\ , \tag{4.1}$$

where $k_{\mu\nu}$ is the tensor introduced above,

$$k_{\mu\nu}^{mn}(\phi)=\sum_{\text{roots}}e_\mu{}^\alpha(\phi)e_\nu{}^{-\alpha}(\phi)(\Lambda_m\cdot\alpha)(\alpha\cdot\Lambda_n)\ .$$

The idea is to express quantities like $H(\phi)$ as functionals of smooth interpolating fields $\phi(x)$,

$$H(\phi)=\int d^Dx\,\mathcal{H}(\phi(x),\partial\phi(x),\dots) \tag{4.2}$$

defining thereby a Hamiltonian density. Since the interpolating fields are supposed to be slowly varying one aims to find the leading terms in an expansion in powers of $\partial\phi,\partial^2\phi$, etc. The lattice sums one meets are usually simple enough that this expansion can be found by inspection. However, it is probably worth pointing out that there is a systematic procedure that employs Fourier expansions. The lattice variables ϕ_n may be represented by Fourier integrals,

$$\phi_n=\int\left[\frac{dk}{2\pi}\right]^D\tilde{\phi}(\underline{k})e^{i\underline{k}\cdot\underline{n}}\ , \tag{4.3}$$

where the wave vectors are integrated over a cell of volume $(2\pi)^D$. The Fourier components $\tilde{\phi}$ are periodic in $\underline{k}$ with period 2π. The Hamiltonian may be regarded as a functional of $\tilde{\phi}$, e.g.,

$$H(\phi)=\sum_N\frac{1}{N!}\int\left[\frac{dk_1}{2\pi}\right]^d\cdots\left[\frac{dk_N}{2\pi}\right]^D H_N(k_1,\dots,k_N)\tilde{\phi}(k_1)\cdots\tilde{\phi}(k_N)\ , \tag{4.4}$$

where the coefficient functions H_N are of course periodic in the wave vectors $\underline{k}_1,\underline{k}_2,\dots,\underline{k}_N$. If ϕ_n is slowly varying then $\tilde{\phi}(\underline{k})$ is nonvanishing only in the neighborhood of $\underline{k}=0(\text{mod}2\pi)$. In this case it is therefore legitimate to expand the coefficient functions in powers of $\underline{k}_1,\dots,k_N$, e.g.,

$$H_N(k_1,\dots,k_N)=(2\pi)^D\delta(\Sigma k)(h_N^{(0)}-\Sigma\underline{k}_j\cdot\underline{k}_l h_N^{(2)}+\cdots)\ .$$

The periodicity in $\underline{k}$ is now irrelevant, and one may substitute

$$(2\pi)^D\delta(\Sigma k)=\int d^Dx\,e^{i\Sigma\underline{k}\cdot\underline{x}}\ .$$

Define the interpolating fields $\phi(x)$ by the continuum Fourier integral,

$$\phi(x)=\int\left[\frac{dk}{2\pi}\right]^D\tilde{\phi}(\underline{k})e^{i\underline{k}\cdot\underline{x}}\ . \tag{4.5}$$

One then obtains $H(\phi)$ in the form (4.2) with

$$\mathcal{H}=\sum_N\frac{1}{N!}\left[h_N^{(0)}\phi(x)^N+\binom{N}{2}h_N^{(2)}\phi(x)^{N-2}(\underline{\partial}\phi)^2+\cdots\right]$$
$$=V(\phi)+Z(\phi)(\underline{\partial}\phi)^2+\cdots\ . \tag{4.6}$$

In the following we shall discard terms containing more than two derivatives of the interpolating fields. Since the ex-

pression (4.1) is bilinear in $\phi_m-\phi_n$ it will be sufficient to keep only the first derivative in this quantity. However, a minor complication arises from the need to keep track of the various translation invariant sublattices. We write

$$\phi_{\underline{n}\nu}=\phi_{\underline{n}}+\xi_{\underline{n}\nu}\ , \tag{4.7}$$

where $\underline{n}$ is a D vector with integer-valued components and $\nu=1,\ldots,f$ labels the sublattice. The variables $\xi_{\underline{n}\nu}$ are constrained to satisfy

$$\sum_{\nu}\xi_{\underline{n}\nu}=0\ , \tag{4.8}$$

and they are small quantities of order $\phi_{\underline{n}}-\phi_{\underline{m}}$. Hence, to leading order,

$$\phi_{\underline{n}\nu}-\phi_{\underline{n}'\nu'}=[(\underline{n}-\underline{n}')\cdot\underline{\partial}\phi(x)+\xi_{\nu}(x)-\xi_{\nu'}(x)]_{x=\underline{n}}\ , \tag{4.9}$$

which is to be substituted into the formula for the Hamiltonian,

$$H(\phi)=\tfrac{1}{2}\sum_{n\nu n'\nu'}J_{\underline{n}\nu,\underline{n}'\nu'}[\Lambda_{\nu}\cdot\Lambda_{\nu'}+\tfrac{1}{2}(\phi_{\underline{n}\nu}-\phi_{\underline{n}'\nu'})^{\sigma}(\phi_{\underline{n}\nu}-\phi_{\underline{n}'\nu'})^{\rho}k^{\nu\nu'}_{\sigma\rho}(\phi_{\underline{n}})+\cdots]\ .$$

Using the translation invariance of $J_{\underline{n}\nu,\underline{n}'\nu'}$, one obtains the density,

$$\mathcal{H}=\tfrac{1}{2}\sum_{\underline{n}\nu\nu'}J_{\underline{n}\nu,\underline{0}\nu'}[\Lambda_{\nu}\cdot\Lambda_{\nu'}+\tfrac{1}{2}(\underline{n}\cdot\underline{\partial}\phi+\xi_{\nu}-\xi_{\nu'})^{\sigma}(\underline{n}\cdot\underline{\partial}\phi+\xi_{\nu}-\xi_{\nu'})^{\rho}k^{\nu\nu'}_{\sigma\rho}(\phi)+\cdots]\ . \tag{4.10}$$

It is generally assumed that the coupling strengths fall off rapidly with distance between sites so that the sum over $\underline{n}$ in this expression contains only a few terms. Defining the moments,

$$\begin{aligned}
&\sum_{\underline{n}}J_{\underline{n}\nu,\underline{0}\nu'}=J_{\nu\nu'}=J_{\nu'\nu}\ ,\\
&\sum_{\underline{n}}n^{i}J_{\underline{n}\nu,\underline{0}\nu'}=J^{i}_{\nu\nu'}=-J^{i}_{\nu'\nu}\ ,\\
&\sum_{\underline{n}}n^{i}n^{j}J_{\underline{n}\nu\cdot\underline{0}\nu'}=J^{ij}_{\nu\nu'}=J^{ij}_{\nu'\nu}\ ,
\end{aligned} \tag{4.11}$$

one obtains

$$\mathcal{H}=\tfrac{1}{2}\sum_{\nu\nu'}\{J_{\nu\nu'}\Lambda_{\nu}\cdot\Lambda_{\nu'}+\tfrac{1}{2}[J^{ij}_{\nu\nu'}\partial_i\phi^{\rho}\partial_j\phi^{\sigma}+2J^{i}_{\nu\nu'}\partial_i\phi^{\rho}(\xi_{\nu}-\xi_{\nu'})^{\sigma}+J_{\nu\nu'}(\xi_{\nu}-\xi_{\nu'})^{\rho}(\xi_{\nu}-\xi_{\nu'})^{\sigma}]k^{\nu\nu'}_{\rho\sigma}(\phi)+\cdots\}\ . \tag{4.12}$$

Although this general result may look rather forbidding, it can simplify in specific applications. For example, if the couplings are isotropic and $J^{i}=0$, $J^{ij}\sim\delta^{ij}$. In the ferromagnetic case there are no sublattices so the variables ξ_{ν} disappear along with the label ν. In the antiferromagnetic case there are usually only two sublattices and $\xi_2=-\xi_1,\Lambda_2=-\Lambda_1$ and, moreover, one usually assumes $J_{11}=J_{22}=0$, etc.

To complete the discussion of the continuum limit, it is necessary to consider the kinetic term,

$$\frac{\hbar}{i}\sum_{\underline{n}\nu}\dot{\phi}^{\mu}_{\underline{n}\nu}A^{j}_{\mu}(\phi_{\underline{n}\nu})\Lambda_{\nu j}\ , \tag{4.13}$$

where A_{μ} is the spin connection on G/H. Substituting the expression (4.7) for $\phi_{\underline{n},\nu}$ and expanding in powers of ξ, one obtains after discarding a total derivative,

$$(\dot{\phi}+\dot{\xi}_{\nu})^{\rho}A_{\rho}(\phi+\xi_{\nu})=\dot{\phi}^{\rho}A_{\rho}(\phi)+\xi^{\rho}_{\nu}(\dot{\phi}^{\lambda}+\tfrac{1}{2}\dot{\xi}^{\lambda}_{\nu})F_{\rho\lambda}+\tfrac{1}{2}\xi^{\rho}_{\nu}\xi^{\sigma}_{\nu}\dot{\phi}^{\lambda}\partial_{\sigma}F_{\rho\lambda}+\cdots\ , \tag{4.14}$$

where $A_{\rho}=A^{j}_{\rho}\Lambda_{\nu j}$ and $F_{\rho\lambda}$ is the corresponding curvature tensor.

In the ferromagnetic case there is nothing further to do. Only the first term in (4.14) is present and the equations of motion will imply $\dot{\phi}/\phi\sim O(k^2)$ to the approximation used in (4.12). In other cases, if we impose the condition,

$$\sum_{\nu}\Lambda_{\nu j}=0\ , \tag{4.15}$$

then the first term in (4.14) is absent and the most important term becomes

$$\frac{\hbar}{i}\sum_{\nu}\xi^{\rho}_{\nu}\dot{\phi}^{\lambda}F^{j}_{\rho\lambda}\Lambda_{\nu j}\ . \tag{4.16}$$

Here the equations of motion imply that both $\dot{\phi}/\phi$ and ξ_{ν} are $O(k)$. It is therefore consistent in this approximation to discard all other terms in (4.14). Notice, in particular, that $\dot{\xi}_{\nu}$ disappears from the Lagrangian. This means that

ξ_ν becomes an algebraic variable and can be eliminated from the dynamics. [If the condition (4.15) is not imposed then ξ_ν becomes a true dynamical variable associated with optical branches in the spectrum, i.e., $\dot{\xi}/\xi$ remains finite in the limit $k \to 0$.]

To conclude, the Lagrangian for the ferromagnetic spin waves is given by

$$L = \int d^D x \left[\frac{\hbar}{i} \partial_t \phi^\mu A_\mu(\phi) - \frac{1}{4} J^{ij} \partial_i \phi^\mu \partial_j \phi^\nu k_{\mu\nu}(\phi) \right] , \tag{4.17}$$

where $A_\mu = A_\mu^j \Lambda_j$ and

$$k_{\mu\nu}(\phi) = \sum_{\text{roots}} e_\mu{}^\alpha e_\nu{}^{-\alpha} (\alpha\Lambda)^2 .$$

The tensor J^{ij} is assumed to be negative definite.

The equations of motion derived from (4.17) are generalizations of the well-known Landau-Lifshitz[15] equations for ferromagnetic spin waves. They are

$$\frac{\hbar}{i} F_{\mu\nu} \partial_t \phi^\nu = \frac{J}{2} k_{\mu\nu} \Delta \phi^\nu ,$$

where we have assumed an isotropic coupling, i.e.,

$$J^{ij} = J\delta^{ij} , \quad J < 0 ,$$

and where the "Laplacian" is given by

$$\Delta \phi^\nu = \partial_i^2 \phi^\nu + \tfrac{1}{2} k^{\nu\rho} (\partial_\lambda k_{\rho\mu} + \partial_\mu k_{\rho\lambda} - \partial_\rho k_{\lambda\mu}) \partial_i \phi^\lambda \partial_i \phi^\mu .$$

It is not difficult to verify that for $G/H = \mathrm{SU}(2)/\mathrm{U}(1)$ these equations assume their familiar form[15]

$$\hbar \partial_t \mathbf{n} = -\frac{J}{2} s \mathbf{n} \times \nabla_i^2 \mathbf{n} ,$$

where s is the spin and $\mathbf{n}$ is a unit vector describing the target space S^2.

For the antiferromagnetic case (with two sublattices),

$$\Lambda_1 = \Lambda = -\Lambda_2 ,$$

$$\xi_1 = \xi = -\xi_2 ,$$

the Lagrangian is given by

$$L = \int d^D x \left[\frac{\hbar}{i} 2\xi^\mu F_{\mu\nu} \partial_t \phi^\nu - (J^{ij} \partial_i \phi^\mu \partial_j \phi^\nu + 4J^i \partial_i \phi^\mu \xi^\nu + 4J \xi^\mu \xi^\nu) k_{\mu\nu}(\phi) \right] , \tag{4.18}$$

where, for simplicity, we have taken $J_{11}^{ij} = J_{22}^{ij} = 0$, and we have suppressed the sublattice indices, writing

$$J_{12}^{ij} = J_{21}^{ij} = J^{ij} , \quad J_{12}^i = -J_{21}^i = J^i ,$$

$$J_{12} = J_{21} = J ,$$

and

$$k_{\mu\nu}^{12} = k_{\mu\nu}^{21} = k_{\mu\nu} = -\sum_{\text{roots}} e_\mu{}^\alpha e_\nu{}^{-\alpha} (\alpha\Lambda)^2 . \tag{4.19}$$

The tensor J^{ij} is assumed to be positive definite. The auxiliary variable ξ^μ can be eliminated from (4.18) by solving the Euler-Lagrange equation $\delta L / \delta \xi^\mu = 0$,

$$\frac{\hbar}{i} F_{\mu\nu} \dot{\phi}^\nu - 2k_{\mu\nu} (J^i \partial_i \phi^\nu + 2J\xi^\nu) = 0 .$$

This gives

$$L = \int d^D x \left[\frac{\hbar^2}{4J} g_{\mu\nu} \partial_t \phi^\mu \partial_t \phi^\nu - \frac{\hbar}{i} \frac{J^i}{J} \partial_i \phi^\mu F_{\mu\nu} \partial_t \phi^\nu - \left[J^{ij} - \frac{J^i J^j}{J} \right] \partial_i \phi^\mu \partial_j \phi^\nu k_{\mu\nu} \right] , \tag{4.20}$$

where the tensor $g_{\mu\nu}(\phi)$ is defined by

$$g_{\mu\nu} = -F_{\mu\lambda} F_{\nu\rho} (k^{-1})^{2\rho} = -\sum_{\text{roots}} e_\mu{}^\alpha e_\nu{}^{-\alpha} . \tag{4.21}$$

[The formulas (4.19) and (3.17) have been used.] The tensor $g_{\mu\nu}$ is G invariant and positive definite.

For symmetric spaces there is a unique G-invariant second-rank symmetric tensor on G/H implying that

$$k_{\mu\nu}(\phi) = v^2 g_{\mu\nu}(\phi)$$

when v is independent of ϕ. If in addition we also have invariance under space rotations, then

$$J^{ij} - \frac{J^i J^j}{J} \sim \delta^{ij}$$

and the theory becomes fully relativistic. This obtains automatically for $D = 1$ for which we recover the well-known two-dimensional σ models targeted on symmetric spaces.

The Lagrangian (4.20) gives a sensible description of antiferromagnetic spin waves if

$$J>0\ ,\quad J^{ij}-\frac{J^iJ^j}{J}>0\ . \tag{4.22}$$

The middle term in (4.20),

$$-\frac{\hbar}{i}\frac{J^i}{J}\partial_t\phi^\mu F_{\mu\nu}\partial_t\phi^\nu \tag{4.23}$$

is of a topological nature. It makes no contribution to the classical equations of motion. Indeed, it can be expressed as a total derivative. The integral of this quantity over a space-time region, M, reduces to a boundary integral,

$$-\frac{\hbar}{i}\frac{J^i}{J}\int_M dt\,d^Dx\,\partial_i\phi^\mu\partial_t\phi^\nu F_{\mu\nu}$$
$$=-\frac{\hbar}{i}\frac{J^i}{J}\int_M dt\,d^Dx[\partial_i(\partial_t\phi^\mu A_\mu)-\partial_t(\partial_i\phi^\mu A_\mu)]\ . \tag{4.24}$$

If M is a compact two-dimensional manifold this integral is a topological invariant—an integer multiple of $2\pi\hbar$ for the two sphere as will be shown in the following section. In general the integral (4.24) has no topological significance. For compact manifolds of higher dimension it has been argued that this term should vanish,[12] at least for smooth field configurations.

It was shown by Haldane, who discovered the term in the case of a one-dimensional antiferromagnetic spin system with nearest-neighbor couplings, that the topological contribution gives an alternating sign in the sum over configurations for half-integer spin systems.[11] In the following we consider some classical aspects of the one-dimensional case.

V. INSTANTONS

The continuum theory that emerges from the one-dimensional antiferromagnetic chain is described by a Lagrangian of the type (4.20). Consider now the Euclidean version of this theory. The Lagrangian density is

$$\mathcal{L}=\frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu+J'k_{\mu\nu}\partial_x\phi^\mu\partial_x\phi^\nu\ , \tag{5.1}$$

where we take $\hbar=1$ and discard the topological term. The tensors $g_{\mu\nu}$ and $k_{\mu\nu}$, given by (4.21) and (4.19), respectively, are positive definite, and we shall assume that J and J' are positive.

There is a simple argument that suggests the possible existence of finite action solutions of the equations of motion, i.e., the instantons. Consider the expressions

$$\frac{1}{4J}g_{\mu\nu}(\partial_t\phi^\mu\pm 2i\sqrt{JJ'}F^\mu{}_\lambda\partial_x\phi^\lambda)(\partial_t\phi^\nu\pm 2i\sqrt{JJ'}F^\nu{}_\rho\partial_x\phi^\rho)=\frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu\pm i\sqrt{J'/J}F_{\nu\lambda}\partial_x\phi^\lambda\partial_t\phi^\nu$$
$$-J'g_{\mu\nu}F^\mu{}_\lambda F^\nu{}_\rho\partial_x\phi^\lambda\partial_x\phi^\rho$$
$$=\frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu\pm i\sqrt{J'/J}F_{\nu\lambda}\partial_x\phi^\lambda\partial_t\phi^\rho+J'k_{\mu\nu}\partial_x\phi^\mu\partial_x\phi^\nu\ , \tag{5.2}$$

where

$$g_{\mu\nu}F^\nu{}_\rho=F_{\mu\rho}=\sum_{\text{roots}}e_\mu{}^\alpha e_\rho{}^{-\alpha}\alpha\cdot\Lambda\ . \tag{5.3}$$

In a basis of real coordinates the components $F_{\mu\nu}$ are pure imaginary. Hence, the expressions (5.2) are real and positive, i.e.,

$$\mathcal{L}=\frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu+J'k_{\mu\nu}\partial_x\phi^\mu\partial_x\phi^\nu$$
$$\geq|\sqrt{J'/J}F_{\mu\nu}\partial_x\phi^\mu\partial_t\phi^\nu|\ .$$

This means that the Euclidean action has a lower bound,

$$\int d^2x\left[\frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu+J'k_{\mu\nu}\partial_x\phi^\mu\partial_x\phi^\nu\right]$$
$$\geq\left|\sqrt{J'/J}\int d^2x\,F_{\mu\nu}\partial_x\phi^\mu\partial_t\phi^\nu\right|\ . \tag{5.4}$$

It remains to determine whether there are configurations for which the integral on the right-hand side is finite.

Suppose that the two-dimensional Euclidean space-time has the topology of a sphere. The coordinate functions, $\phi^\mu(x)$, define the image of this sphere in the manifold, G/H. It is not difficult to see that such maps have a topological classification. The manifold is generally covered by more than one coordinate patch. If the coordinates in two patches that intersect are associated with the group elements $L_\phi^{(1)}$ and $L_\phi^{(2)}$, as described above in Sec. II, then at points in the overlap there is a relation,

$$L_\phi^{(2)}=L_\phi^{(1)}h^{(1,2)}(\phi)\ , \tag{5.5}$$

where $h^{(1,2)}\in H$. Only the Abelian part of $h^{(1,2)}$ is of interest because this is what determines the relation between $A^j(\phi)$ in the two patches. With

$$h^{(1,2)}(\phi)=e^{i\psi^j(\phi)H_j} \tag{5.6}$$

it follows from the definition (2.12) that

$$A^j_{(2)}=A^j_{(1)}-i\,d\psi^j\ . \tag{5.7}$$

If the overlap between the two patches is not simply connected then the angles $\psi^j(\phi)$ can be multiple valued, i.e.,

$$\Delta\psi^j=\oint d\psi^j\neq 0\ ,$$

where the integral is taken around a closed path in the overlap. However, it is essential that the group element $h^{(1,2)}$ be single valued and this implies

$$\Delta\psi^j\Lambda_j \in 2\pi Z \ . \tag{5.8}$$

On the other hand, the flux of $F^j\Lambda_j$ associated with a compact two space must reduce to a sum of such terms. With the two sphere suppose that the northern hemisphere maps into patch 1, while the southern maps into patch 2. Then the flux is given by

$$\begin{aligned}\int F^j\Lambda_j &= \int dA^j\Lambda_j \\ &= \oint (A^j_{(1)} - A^j_{(2)})\Lambda_j = i\Delta\psi^j\Lambda_j \ ,\end{aligned} \tag{5.9}$$

where the contour runs around the image of the equator. The flux is an integer multiple of $2\pi i$. This integer, N, serves to classify the configurations and, in each class the Euclidean action is bounded below by $|2\pi N\sqrt{J'/J}\,|$.

To saturate the bound one needs to solve the first-order differential equations,

$$g_{\mu\nu}\partial_t\phi^\nu + 2i\sqrt{JJ'}F_{\mu\nu}\partial_x\phi^\nu = 0$$

or, more explicitly,

$$(\partial_t\phi^\mu - 2i\sqrt{JJ'}\alpha\cdot\Lambda\partial_x\phi^\mu)e_\mu{}^\alpha = 0 \ . \tag{5.10}$$

The solutions of these equations comprise the instantons. If G/H is a complex symmetric space it can be shown that Eq. (5.10) reduces to the well-studied case[16]

$$\partial_{\bar z}\zeta^i = 0 \ ,$$

where z is a complex coordinate in the two-dimensional Euclidean space-time and $(\zeta^i,\bar\zeta^i)$ is a set of complex coordinates on the target space G/H.

Returning to Haldane's observation, it will be seen that our result (4.20) appears to differ in some respects. The Euclidean version of (4.20) with topological term included is

$$\begin{aligned}\mathcal{L} = \frac{1}{4J}g_{\mu\nu}\partial_t\phi^\mu\partial_t\phi^\nu + \frac{J^x}{J}F_{\mu\nu}\partial_x\phi^\mu\partial_t\phi^\nu \\ + \left[J^{xx} - \frac{J^xJ^x}{J}\right]k_{\mu\nu}\partial_x\phi^\mu\partial_x\phi^\nu \ ,\end{aligned} \tag{5.11}$$

where we have again set $\hbar = 1$. The coefficients J, J^x and J^{xx} are given in terms of the spin couplings, $J_{n\sigma;n'\sigma'}$, by the lattice sums (4.11). In particular,

$$J = \sum_n J_{n,1;o,2} \quad \text{and} \quad J^x = \sum_n nJ_{n,1;o,2} \ . \tag{5.12}$$

Due to the presence of the topological term the Euclidean action contains the imaginary contribution

$$2\pi i N J^x/J \ ,$$

where N is the instanton number. Haldane's result, obtained for SU(2) spin models with nearest-neighbor couplings, is simply $i\pi N$. In fact, this result does agree with ours, since it can be shown quite generally that $J^x = J/2$. The relevant coupling parameters are

$$\begin{aligned}J_{n,1;n,2} &= J_{n+1,1;n2} = K_1 \ , \\ J_{n-1,1;n2} &= J_{n+2,1;n2} = K_3 \ , \\ J_{n-2,1;n2} &= J_{n+3,1;n2} = K_5 \ ,\end{aligned} \tag{5.13}$$

etc., corresponding to nearest, third-nearest, fifth-nearest, etc., neighbors on the lattice. Substituting into (5.12) gives

$$\begin{aligned}J &= 2K_1 + 2K_3 + 2K_5 + \cdots \ , \\ J^x &= K_1 + K_3 + K_5 + \cdots \ ,\end{aligned} \tag{5.14}$$

which confirms Haldane's result in the more general context.

To summarize, the Euclidean path integral for the generalized antiferromagnetic spin chains includes the sign factor, $e^{i\pi N}$, where N is an integer defined by the flux integral,

$$N = \frac{1}{2\pi}\int F^j\Lambda_j = \frac{1}{2\pi}\int F_j\Lambda^j = N_j\Lambda^j \ . \tag{5.15}$$

The weights $\Lambda^j = g^{jk}\Lambda_k$ are integers for any finite-dimensional unitary representation of G. The coefficients N_j are integers that characterize the configuration $\phi^\mu(x)$. Models in which the Λ^j are all even will not be sensitive to the classes of configurations—like the integer spin SU(2) models. Other models, in which one or more of the Λ^j are odd, will be sensitive—they must generalize, in various ways, the half-integer spin SU(2) models.

VI. OPERATOR REALIZATION FOR SU($N+1$)

To resolve the factor ordering ambiguities in the quantized theory it is necessary to construct a realization of the generators $Q_\alpha(\phi)$ in terms of the dynamical variables ϕ_n. The details of such a realization will depend on the nature of the target manifold G/H and the quantum numbers of the reference state $|\Lambda\rangle$. We are not able to give a general solution but we can illustrate a method that works in at least some cases. Here we discuss the realization of SU($N+1$) in terms of operators associated with the coordinates of CP^N.

The algebra of SU($N+1$) is spanned by the $N(N+2)$ generators, $Q_A{}^B$, $A,B = 1,2,\ldots,N+1$, which satisfy the commutation rules

$$[Q_A{}^B, Q_C{}^D] = \delta^D_A Q_C{}^B - \delta^B_C Q_A{}^D \tag{6.1}$$

and the constraint $Q_A{}^A = 0$. The fundamental representation of this algebra is generated by the $(N+1)$-dimensional traceless Hermitian matrices,

$$(Q_A{}^B)_C{}^D = \delta^D_A\delta^B_C - \frac{1}{N+1}\delta^B_A\delta^D_C \ . \tag{6.2}$$

We shall consider cosets of SU($N+1$) with respect to the subgroup SU(N)×U(1) generated by the N^2 operators, $Q_i{}^j$, $i,j = 1,\ldots,N$. The finite transformations, L_ϕ, that represent the cosets can be chosen such that, in the fundamental representation,

$$(L_\phi)_C{}^D = \begin{bmatrix} \sqrt{1-\phi\phi^\dagger} & \phi \\ -\phi^\dagger & \sqrt{1-\phi^\dagger\phi} \end{bmatrix} , \tag{6.3}$$

where ϕ is an N-component column vector. The coordinates ϕ_i on the manifold CP^N belong to the fundamental representation of SU(N). It is not difficult to verify that the matrix (6.3) is both unitary and unimodular.

According to the general discussion of Sec. II, the reference state $|\Lambda\rangle$, from which the coherent states are generated, must be invariant up to a phase under the action of the stability group SU(N)×U(1). In particular, it must be a singlet of SU(N). We are therefore restricted to those representations of SU($N+1$) that contain such a singlet. In terms of Young tableaux these representations (n_1,n_2) are characterized by n_1 columns with N boxes and n_2 columns with 1 box. Each of these representations contains just one singlet of SU(N). We shall find that the realizations to be obtained depend on n_1-n_2, only.

The Cartan subalgebra of SU($N+1$) is spanned by the N elements, $Q_1{}^1, Q_2{}^2, \ldots, Q_N{}^N$, whose eigenvalues define the components of the weight vectors

$$Q_j{}^j|\Lambda\rangle=|\Lambda\rangle\Lambda_j\ ,\quad j=1,\ldots,N\ .$$

Since the reference state is a singlet of SU(N) these components are all equal

$$\Lambda_1=\Lambda_2=\cdots=\Lambda\ . \tag{6.4}$$

It is convenient to define the hypercharge operator

$$Y=Q_{N+1}{}^{N+1}=-\sum_1^N Q_j{}^j \tag{6.5}$$

to stand for the SU(N) singlet generator. In the fundamental representation it is a diagonal matrix,

$$Y_C{}^D=\begin{bmatrix} -1/(N+1) & & & \\ & \ddots & & \\ & & -1/(N+1) & \\ & & & N/(N+1) \end{bmatrix}. \tag{6.6}$$

The value of the hypercharge in the reference state is given by

$$Y|\Lambda\rangle=-\sum_j Q_j{}^j|\Lambda\rangle=-N\Lambda|\Lambda\rangle\ .$$

On the other hand, since this state is the singlet component of a representation contained in the direct product of n_2 fundamental representations and n_1 conjugates, it follows that

$$Y=(n_2-n_1)\frac{N}{N+1}$$

or

$$\Lambda=\frac{n_1-n_2}{N+1}\ . \tag{6.7}$$

The U(1)-component of the spin connection of CP^N is defined by

$$L_\phi^{-1}dL_\phi=iAY+\cdots\ .$$

With the parametrization chosen in (6.3) this gives

$$A=\frac{1}{i}\frac{N+1}{2N}(\phi^\dagger d\phi-d\phi^\dagger\phi)\ . \tag{6.8}$$

The overlap between neighboring coherent states is, therefore,

$$\begin{aligned}\langle\phi+d\phi|\phi\rangle&=\langle\Lambda|L^{-1}_{\phi+d\phi}L_\phi|\Lambda\rangle\\&=1-iA\langle\Lambda|Y|\Lambda\rangle\\&=1+\tfrac{1}{2}(N+1)\Lambda(\phi^\dagger d\phi-d\phi^\dagger\phi)\ ,\end{aligned}$$

where Λ is given by (6.7). Hence, the Lagrangian (2.18) in this case takes the form

$$L=\frac{\hbar}{2i}(N+1)\Lambda(\phi^\dagger\partial_t\phi-\partial_t\phi^\dagger\phi)-H(\phi,\phi^\dagger)\ . \tag{6.9}$$

Canonical quantization gives the commutation rules

$$\begin{aligned}&[\phi_i,\phi_j]=0\ ,\\&[\phi_i,\phi_j^\dagger]=\frac{-1}{(N+1)\Lambda}\delta_{ij}\ ,\\&[\phi_i^\dagger,\phi_j^\dagger]=0\ ,\end{aligned} \tag{6.10}$$

where $(N+1)\Lambda$ is an integer.

To construct the generators out of these operators it is helpful to consider firstly their coherent state expectation values,

$$\begin{aligned}\langle\phi|Q_A{}^B|\phi\rangle&=\langle\Lambda|L_\phi^{-1}Q_A{}^BL_\phi|\Lambda\rangle\\&=(L_\phi)_A{}^C\langle\Lambda|Q_C{}^D|\Lambda\rangle(L_\phi^{-1})_D{}^B\\&=-(N+1)\Lambda(L_\phi YL_\phi^{-1})_A{}^B\ ,\end{aligned}$$

where Y is the hypercharge matrix in the fundamental representation (6.6). With the matrices $(L_\phi)_A{}^B$ given by (6.3) one obtains

$$\langle\phi|Q_A{}^B|\phi\rangle=\begin{bmatrix}\Lambda-(N+1)\Lambda\phi\phi^\dagger & -(N+1)\Lambda\sqrt{1-\phi^\dagger\phi}\phi\\ -(N+1)\Lambda\sqrt{1-\phi^\dagger\phi}\phi^\dagger & -N\Lambda+(N+1)\Lambda\phi^\dagger\phi\end{bmatrix}, \tag{6.11}$$

using the elementary identities

$$\sqrt{1-\phi\phi^\dagger}\phi=\phi\sqrt{1-\phi^\dagger\phi}$$

and

$$\phi^\dagger\sqrt{1-\phi\phi^\dagger}=\sqrt{1-\phi^\dagger\phi}\phi^\dagger\ .$$

The classical expressions on the right-hand side of (6.11) suggest the following operator realization for $\Lambda<0$,

$$\begin{aligned}
&Q_i{}^j(\phi)=\Lambda\delta_i^j-(N+1)\Lambda\phi_j^\dagger\phi_i=(\Lambda-1)\delta_i^j-(N+1)\Lambda\phi_i\phi_j^\dagger\ ,\\
&Q_i{}^{N+1}(\phi)=-(N+1)\Lambda\sqrt{1-\phi^\dagger\phi}\,\phi_i=-(N+1)\Lambda\phi_i\sqrt{1-1/(N+1)\Lambda-\phi^\dagger\phi}\ ,\\
&Q_{N+1}{}^j(\phi)=-(N+1)\Lambda\phi_j^\dagger\sqrt{1-\phi^\dagger\phi}=-(N+1)\Lambda\sqrt{1-1/(N+1)\Lambda-\phi^\dagger\phi}\,\phi_j^\dagger\ ,\\
&Q_{N+1}{}^{N+1}(\phi)=-N\Lambda+(N+1)\Lambda\phi^\dagger\phi\ .
\end{aligned}\tag{6.12}$$

It is easy to verify, using the commutation rules (6.10) that these operators satisfy the algebra of SU($N+1$). The reference state is annihilated by ϕ_i. The basis for a finite dimensional unitary representation of SU($N+1$) is given by

$$|\Lambda\rangle,\phi_i^\dagger|\Lambda\rangle,\phi_{i_1}^\dagger\phi_{i_2}^\dagger|\Lambda\rangle,\ldots,\phi_{i_1}^\dagger\ldots\phi_{i_p}^\dagger|\Lambda\rangle\ ,\tag{6.13}$$

where $p=-(N+1)\Lambda=n_2-n_1$ is a positive integer. The rest of the Fock space, spanned by vectors with $p+1,p+2,\ldots$creation operators, carries a nonunitary infinite dimensional representation of SU($N+1$).

Similarly, for $\Lambda>0$ the operator realization is given by

$$\begin{aligned}
&Q_i{}^j(\phi)=\Lambda\delta_i^j-(N+1)\Lambda\phi_i\phi_j^\dagger=(\Lambda+1)\delta_i^j-(N+1)\Lambda\phi_j^\dagger\phi_i\ ,\\
&Q_i{}^{N+1}(\phi)=-(N+1)\Lambda\phi_i\sqrt{1+N/(N+1)\Lambda-\phi^\dagger\phi}=-(N+1)\Lambda\sqrt{1+1/\Lambda-\phi^\dagger\phi}\,\phi_i,\\
&Q_{N+1}{}^j(\phi)=-(N+1)\Lambda\sqrt{1+N/(N+1)\Lambda-\phi^\dagger\phi}\,\phi_i^\dagger=-(N+1)\Lambda\phi_i^\dagger\sqrt{1+1/\Lambda-\phi^\dagger\phi}\ ,\\
&Q_{N+1}{}^{N+1}(\phi)=-(\Lambda+1)N+(N+1)\Lambda\phi^\dagger\phi\ .
\end{aligned}\tag{6.14}$$

Here the reference is annihilated by $\phi_i^\dagger$ and the basis for a finite dimensional unitary representation is given by

$$|\Lambda\rangle,\phi_i|\Lambda\rangle,\phi_{i_1}\phi_{i_2}|\Lambda\rangle,\ldots,\phi_{i_1}\ldots\phi_{i_p}|\Lambda\rangle\ ,\tag{6.15}$$

where $p=(N+1)\Lambda=n_1-n_2$.

It appears from the structure of the basis sets, (6.13) and (6.15) that the operators (6.12) and (6.14) realize, respectively, the unitary representations $(0,n_2-n_1)$ and $(n_1-n_2,0)$ but not the general case (n_1,n_2). Our approach clearly leaves something to be desired. This should not be a surprise, since we began with the assumption that commutators should be ignored in the first approximation—an assumption that is appropriate to the correspondence theory limit. The resulting formula should, therefore, be read as corrections to this limiting case valid, we expect, for large n_1 and n_2 such that

$$|n_1-n_2|\ll n_1+n_2\ .$$

We believe that it should be possible to construct exact expressions from ϕ and $\phi^\dagger$ that realize the general case (n_1,n_2) and that reduce to the forms (6.12) and (6.14) when higher powers of $|n_1-n_2|/(n_1+n_2)$ are neglected. Probably these expressions will have branch point singularities at $\phi=0$. In support of the conjecture we have examined a Holstein-Primakoff–type realization for the SU(2) case, which depends explicitly on the spin s and helicity λ and which reduces to the standard Holstein-Primakoff form when $\lambda=\pm s$. This realization is discussed in the Appendix.

Finally, in applying these formulas to an SU($N+1$) lattice model one must construct the Hamiltonian operator, $H(\phi,\phi^\dagger)$ and set up the Feynman rules. Typically, the free propagator $\langle T\phi_n\phi_m^\dagger\rangle$ will have the Fourier representation

$$G(\omega,k)=\frac{1}{(n_1-n_2)\hbar\omega-H(k)}$$

and the higher-order contributions will be suppressed by powers of $(n_1-n_2)^{-1}$. Corrections to the classical theory are obtained in the usual way, as a loop expansion. Of course, at any finite order the approximate charge operators will give rise to transitions into the unphysical (nonunitary) part of the representation. This defect is endemic to the semiclassical treatment of spin waves.

VII. OUTLOOK

In this paper we have begun the investigation of a class of spin systems based on an arbitrary Lie group. The purpose of this study is to develop a better understanding of the role of symmetry in the behavior of spin systems in general. We believe that it may eventually be possible to generalize some of the rigorous work that has traditionally been concentrated on the SU(2) systems, particularly that which concerns the nature of the ground state: the question of long-range order and its dependence on parameters such as spin, coupling strength and the dimensionality of space. At this point we have not addressed such questions. The work described in this paper is of a preliminary nature, being concerned with kinematics and only the simplest approximations: the correspondence-theory limit and the naive long-wavelength limit. Even so, we believe that the results have some interest.

First, we have shown that the classical theory that emerges in the correspondence-theory limit is associated with a flag manifold or one of its submanifolds. These manifolds are not, in general, symmetric spaces although they are homogeneous and can be analyzed by group theoretic methods. At least the preliminary features of

the semiclassical theory can be discerned in that it is possible to identify dynamical variables and formulate a canonical quantization program. The problem of factor ordering, however, has been only partially resolved. To express interesting Hamiltonians as well-defined self-adjoint and group-invariant operators it will probably be necessary to obtain more general versions of the Holstein-Primakoff realization than we have found. Our version, discussed in Sec. VI, is able to realize only a limited class of representations, those for which the weights are not degenerate. These representations are characterized by a single quantum number, n_1-n_2 in the case of $SU(N+1)$. One of our motivations for going beyond SU(2) was to have a richer variety of quantum numbers on which interesting physical quantities might depend. This remains to be achieved.

Second, we have shown that the long-wavelength properties of the classical theory are described by generalized nonrelativistic σ models in which the fields take their values on a coset space G/H in which H includes the maximal torus. The Lagrangian is of first order in the time derivative but, depending on the nature of the ground-state ordering, it may be equivalent to a second-order theory. Such second-order systems resemble the relativistic σ models except that the space and time derivatives are coupled to independent tensors, $k_{\mu\nu}$ and $g_{\mu\nu}$, respectively. The flag manifold generally admits more than one invariant tensor of rank 2. This is to be contrasted with the homogeneous symmetric spaces, which admit only one: σ models on symmetric spaces such as CP^N are inevitably relativistic.

A question that calls for investigation is the long-wavelength behavior of the quantized theory. At present it is not clear whether the factor ordering ambiguities are significant in the continuum limit. If they are not, then it should be possible to set up renormalization-group equations and search for fixed points. In particular, it would be interesting, in the case of second-order σ models to see whether the tensors $k_{\mu\nu}$ and $g_{\mu\nu}$ evolve towards the relativistic form, $k_{\mu\nu}=c^2 g_{\mu\nu}$, at the fixed point. This idea is pursued in a separate paper.[17]

Finally, for generalized spin systems in one space dimension we have shown that the Haldane topological term emerges in the continuum limit of the antiferromagnetic chain. As in the the SU(2) case it appears that representations for which the contravariant weight components, Λ^j, are odd integers are distinguished. They contribute an alternating sign to the path integrals. In higher than one space dimension, it should be interesting to examine the effect of the singular "hedghogs" on the phase structure of our generalized models, along the lines developed by Read and Sachdev.[13] This may shed some light on possible physical applications of the spin system studied in this paper.

APPENDIX: THE REALIZATION OF SU(2)

A generalization of the Holstein-Primakoff realization can be obtained by applying the method discussed in Sec. II. The structures here are sufficiently simple that all formulas can be exhibited in detail and so provide a useful illustration of the method. In particular, we shall be able to see how the Holstein-Primakoff formulas emerge as a special case.

Consider the $(2s+1)$-dimensional representation of SU(2) defined by

$$\begin{aligned} &J_3|m\rangle=|m\rangle m\ , \\ &J_+|m\rangle=|m+1\rangle\sqrt{s-m)(s+m+1)}\ , \\ &J_-|m\rangle=|m-1\rangle\sqrt{(s+m)(s-m+1)}\ , \end{aligned} \tag{A1}$$

where $m=-s,-s+1,\ldots,s$ and $\langle m|m'\rangle=\delta_{mm'}$. The coherent states are defined by

$$|\theta\varphi\rangle=L_{\theta\varphi}|\lambda\rangle=e^{-i\varphi J_3}e^{-\theta J_2}e^{i\varphi J_3}|\lambda\rangle \tag{A2}$$

where $0\le\theta<\pi$, $0\le\varphi\le 2\pi$ and $|\lambda\rangle$ is one of the eigenstates of J_3. The expectation value of $\underline{J}$ is a three-vector of length $|\lambda|$,

$$\begin{aligned} &\langle\theta\varphi|J_\pm|\theta\varphi\rangle=\lambda\sin\theta e^{\pm i\varphi} \\ &\langle\theta\varphi|J_3|\theta\varphi\rangle=\lambda\cos\theta\ . \end{aligned} \tag{A3}$$

Applying the methods of Sec. II, computing the infinitesimal overlap of coherent states and expressing the transition amplitudes in path integral notation one arrives at the classical Lagrangian

$$L=\hbar\lambda(\cos\theta-1)\partial_t\varphi-H(\theta,\varphi)\ , \tag{A4}$$

where $(\cos\theta-1)d\varphi$ is the spin connection on the two sphere in polar coordinates. The canonical variables are

$$\varphi \quad\text{and}\quad \pi=\hbar\lambda(\cos\theta-1)\ .$$

In the quantized theory they are represented by operators subject to the commutation rule

$$[\pi,\varphi]=\frac{\hbar}{i}$$

or, more precisely, since states should be represented by periodic functions of φ,

$$e^{i\varphi}\cos\theta e^{i\varphi}=\cos\theta-\frac{1}{\lambda}\ . \tag{A5}$$

In the basis which diagonalizes φ the eigenstates of J_3 are represented by

$$\langle\varphi|m\rangle=e^{im\varphi}\ . \tag{A6}$$

The canonical momentum can be represented by the differential operator

$$\pi=-\hbar(i\partial_\varphi+\lambda) \tag{A7}$$

which gives $\lambda\cos\theta=-i\partial_\varphi$. With this choice one finds, using (A1) and (A6),

$$\begin{aligned} &J_3=\lambda\cos\theta\ , \\ &J_+=e^{i\varphi}\sqrt{(s-\lambda\cos\theta)(s+\lambda\cos\theta+1)} \\ &\quad=\sqrt{(s-\lambda\cos\theta+1)(s+\lambda\cos\theta)}e^{-i\varphi}\ . \end{aligned} \tag{A8}$$

It is straightforward to verify that the operators (A8)

satisfy the commutation rules of SU(2) as well as the constraint

$$\tfrac{1}{2}(J_+J_-+J_-J_+)+J_3^2=s(s+1)\ . \tag{A9}$$

It is clear from (A5) that the operators φ and θ become commutative in the limit $|\lambda|\to\infty$. Also the large quantum number behavior of the operators (A8) approximates the classical formulas (A3). For example, under the assumptions

$$s^2-\lambda^2\lesssim O(|\lambda|) \quad \text{and} \quad |\lambda|\sin^2\theta\gg 1 \tag{A10}$$

as s and $|\lambda|$ become large, one finds the expansion

$$J_+\simeq\lambda e^{i\varphi}\sin\theta\left[1+\frac{s(s+1)-\lambda^2-\lambda\cos\theta}{2\lambda^2\sin^2\theta}-\frac{1}{8}\left[\frac{s(s+1)-\lambda^2-\lambda\cos\theta}{\lambda^2\sin^2\theta}\right]^2+\cdots\right]. \tag{A11}$$

To recover the usual form of the Holstein-Primakoff realization define the operator

$$\phi=e^{i\varphi}\sqrt{(1-\cos\theta)/2}=\sqrt{(1-1/\lambda-\cos\theta)/2}e^{i\varphi} \tag{A12}$$

and its Hermitian conjugate $\phi^\dagger$. The commutation rule becomes

$$[\phi,\phi^\dagger]=-\frac{1}{2\lambda}\ , \tag{A13}$$

which is the SU(2) version of (6.10). The definition (A12) can be inverted to give

$$\cos\theta=1-2\phi^\dagger\phi \tag{A14}$$

and

$$e^{-i\varphi}=\phi\frac{1}{\sqrt{\phi^\dagger\phi}}=\frac{1}{\sqrt{\phi\phi^\dagger}}\phi\ .$$

Hence, substituting into (A8),

$$\begin{aligned} J_3&=\lambda-2\lambda\phi^\dagger\phi\ ,\\ J_+&=\phi^\dagger\sqrt{(s+\lambda-2\lambda\phi\phi^\dagger)(s-\lambda+2\lambda\phi^\dagger\phi)/\phi\phi^\dagger}\\ &=\sqrt{(s+1-\lambda+2\lambda\phi^\dagger\phi)(s+\lambda-2\lambda\phi^\dagger\phi)/\phi^\dagger\phi\phi^\dagger}\\ J_-&=J_+^\dagger\ . \end{aligned} \tag{A15}$$

The singularity at $\phi=0$ in these formulas is eliminated by choosing $s=\pm\lambda$. For example, with $\lambda=-s$ (A15) reduces to the familiar form

$$\begin{aligned} J_3&=-s+2s\phi^\dagger\phi\ ,\\ J_+&=2s\phi^\dagger\sqrt{1-\phi^\dagger\phi}=2s\sqrt{1+1/2s-\phi^\dagger\phi}\phi^\dagger \end{aligned} \tag{A16}$$

with $[\phi,\phi^\dagger]=1/2s$. This is the SU(2) version of the realization (6.12). On the basis of this simple example we are encouraged to believe that it may be possible to generalize the operators (6.12) and (6.14) to yield the representation (n_1,n_2) of SU$(N+1)$.

[1]For reviews, see, for example, I. Affleck, in *Strings, Fields and Critical Phenomena*, Les Houches Summer School, 1988, Session XLIX, edited by E. Brezin and J. Zinn-Justin (North-Holland, Amsterdam, 1990); E. Fradkin, *Field Theories of Condensed Matter Systems* (Addison-Wesley, Redwood City, CA, 1991); E. Manousakis, Rev. Mod. Phys. **63**, 1 (1991).

[2]I. Affleck, Nucl. Phys. **B257**, 397 (1985); **B265**, 409 (1986).

[3]H. A. Bethe, Z. Phys. **71**, 205 (1931); R. L. Orbach, Phys. Rev. **112**, 309 (1958).

[4]C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).

[5]F. D. M. Haldane, Phys. Lett. **93A**, 464 (1983); Phys. Rev. Lett. **50**, 1153 (1983); see also Ref. 2.

[6]See, for example, the discussion in Sec. I of S. Chakravarty, B. I. Halperin, and D. R. Nelson, Phys. Rev. B **39**, 2344 (1989).

[7]See, for example, the review by Manousakis, Ref. 1.

[8]N. Read and S. Sachdev, Nucl. Phys. **B316**, 609 (1989).

[9]I. Affleck, Nucl. Phys. **B305**, 582 (1988).

[10]F. D. M. Haldane, Phys. Rev. Lett. **57**, 1488 (1986); see also A. Perelomov, *Generalized Coherent States and their Applications* (Springer, Berlin, 1986).

[11]F. D. M. Haldane, J. Appl. Phys. **57**, 3359 (1985).

[12]E. Fradkin and M. Stone, Phys. Rev. B **38**, 7215 (1988); F. D. M. Haldane, Phys. Rev. Lett. **61**, 1029 (1988); X. G. Wen and A. Zee, *ibid.* **61**, 1025 (1988); D. Dombre and N. Read, Phys. Rev. B **38**, 7181 (1988).

[13]N. Read and S. Sachdev, Phys. Rev. B **42**, 4568 (1990).

[14]T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).

[15]E. M. Lifshitz and L. P. Pitaevskii, *Statistical Physics, Part 2* (Pergamon, Oxford, 1980).

[16]A. Belavin and A. Polyakov, Zh. Eksp. Teor. Fiz. **22**, 503 (1975) [JETP Lett. **22**, 245 (1975)]; A. D'Adda, P. di Vecchia, and M. Lüscher, Nucl. Phys. **B146**, 63 (1978); M. Lüscher, *ibid.* **B135**, 1 (1978); H. Eichenherr, *ibid.* **B146**, 250 (1978); **B155**, 544 (1979).

[7]S. Randjbar-Daemi and J. Strathdee, Int. J. Mod. Phys. A (to be published).

Bio Data

Born
29 January 1926 in Jhang, Pakistan

Education

Was educated at Panjab University, St. John's College, Cambridge and Cavendish Laboratory, Cambridge (Ph.D. 1952). All degrees conferred with highest honours.

Appointments

Professor, Government College (Lahore)	(1951–1954)
Head of the Mathematics Department, Panjab University (Lahore)	(1951–1954)
Lecturer, Cambridge University (Cambridge)	(1954–1956)
Professor of Theoretical Physics, London University, Imperial College (London)	(1957–1993)
Founder and Director, International Centre for Theoretical Physics (Trieste)	(1964–1993)
Elected Fellow, St. John's College (Cambridge)	(1951–1956)
Member, Institute of Advanced Study (Princeton)	(1951)
Elected, Honorary Life Fellow, St. John's College (Cambridge)	(1971)
Awarded Honorary Professorship, Peking University (Peking)	(1987)
Emeritus Scientist of Centro Brasileiro de Pesquisas Fisicas – CBPF (Rio de Janeiro)	(1992)

Emeritus Professor of Theoretical Physics, London University Imperial College (London)	(1993–)
President, International Centre for Theoretical Physics (Trieste)	(1994–)

United Nations Assignments

Scientific Secretary, Geneva Conferences on Peaceful Uses of Atomic Energy	(1955, 1958)
Elected Member of the Board of Governors, IAEA, Vienna	(1962–1963)
Member, United Nations Advisory Committee on Science and Technology	(1964–1975)
Elected Chairman, United Nations Advisory Committee on Science and Technology	(1971–1972)
Member, United Nations Panel and Foundation Committee for United Nations University	(1970–1973)
Member, United Nations University Advisory Committee	(1981–1983)
Member, Council, University for Peace (Costa Rica)	(1981–1986)
Elected Chairman, UNESCO Advisory Panel on Science, Technology and Society	(1981)

Other Assignments

Member, Scientific Council of SIPRI. First President of the Third World Academy of Science. First President of the Third World Network of Scientific Organisations. Member of the Board of Directors of the Beijir Institute of the Royal Swedish Academy of Sciences. Member of the South Commission. Has also served as Vice President of IUPAP and as a member of the CERN Scientific Policy Committee.

Awards for Contributions to Physics

Between 1958 and 1993, received 18 awards in 10 countries, including the Nobel Prize for Physics in 1979 and the Copley Medal of the Royal Society in 1990.

Other Awards

Has received nine medals for contributions towards peace and promotion of international scientific collaboration between 1968 and 1993.

Academies, Societies and Orders

Has been elected member of prestigious Academies and Societies in 24 Third World and industrialised countries. Is also an honorary member of five important national Orders on three continents as follows:

Order of Nishan-e-Imtiaz (Pakistan)	(1979)
Order of Andres Bello (Venezuela)	(1980)
Order of Istiqlal (Jordan)	(1980)
Cavaliere di Gran Croce dell'Ordine al Merito della Repubblica Italiana	(1980)
Honorary Knight Commander of the Order of the British Empire	(1989)

D.Sc. Honoris Causa

Has received 45 Doctor Honoris Causa awards in 28 countries.

Published Papers

Has published 275 scientific papers on physics of elementary particles, several books, and numerous papers and articles on scientific and educational policies for developing countries.

This page is intentionally left blank

List of Publications in Chronological Order

1943

[1] *A Problem of Ramanujam,* Math. Student **XI**, 1–2.

1950

[2] *Differential Identities in Three-Field Renormalization Problem,* Phys. Rev. **79**, 910–911.

1951

[3] *Overlapping Divergences and the S-Matrix,* Phys. Rev. **82**, 217–227.

[4] (with P. T. Matthews) *The Renormalization of Meson Theories,* Rev. Mod. Phys. **23**, 311–314.

[5] *Divergent Integrals in Renormalizable Field Theories,* Phys. Rev. **84**, 426–431.

1952

[6] *Renormalization of Scalar Electrodynamics Using b-Formalism,* Pakistan J. Sci. **4**, No. 1.

[7] *Recent Advances in Nuclear Theory and Experiment,* Pakistan J. Sci. **4**, No. 1.

[9] (with P. T. Matthews) *The Intermediate Coupling Theory of the Pseudoscalar Meson-Nucleon Interaction,* Phys. Rev. **86**, 715–744.

[8] *Renormalized S-Matrix for Scalar Electrodynamics,* Phys. Rev. **86**, 731–744.

1953

[10] *Fredholm Solutions of Partial Integral Equations,* Proc. Cambridge Phil. Soc. **49**, 213–217.

[11] *Cosmological Theory,* Pakistan J. Sci. **5**, No. 2.

[12] (with P. T. Matthews) *Fredholm Theory of Scattering in a Given Time-Dependent Field,* Phys. Rev. **90**, 690–695.

[13] *The Field Theory of Superconductivity,* Prog. Theor. Phys. **9**, 550–554.

[14] *Modified Propagation Function in Perturbation Theory,* Proc. Cambridge Phil. Soc. **49**, 638–641.

[15] (with P. T. Matthews) *Covariant Fock Equations,* Proc. R. Soc. (London) **A221**, 128–134.

1954

[16] (with P. T. Matthews) *The Green's Functions of Quantized Fields*, Nuovo Cimento **12**, 563–565.

[16a] (with P. T. Matthews) *Renormalization*, Phys. Rev. **94**, 185–191.

1955

[17] (with J. Hamilton) *The Bound State Representation*, Proc. Cambridge Phil. Soc. **51**, 103–112.

[18] (with N. Kemmer) *On the Relativistic Equation for Scattering*, Proc. R. Soc. (London) **A230**, 266–271.

[19] (with P. T. Matthews) *On the Spin of the θ^0-Meson*, Phil. Mag.**46**, 150–154.

[20] (with P. T. Matthews) *Propagators of Quantized Field*, Nuovo Cimento **2**, 120–134.

[21] (with J. C. Polkinghorne) *On the Classification of Fundamental Particles*, Nuovo Cimento **2**, 685–690.

1956

[22] *On Generalized Dispersion Relations I*, Nuovo Cimento **3**, 424–429.

[23] *Opening of the Tait Institute of Mathematical Physics, Edinburgh*, Nucl. Phys. **1**, 366–378.

[24] (with W. Gilbert) *On Generalized Dispersion Relations II*, Nuovo Cimento **3**, 607–611.

[25] *Dispersion Theory*, CERN Symp., Vol. 2.

[25a] *Lectures on Dispersion Theory*, University of Rochester.

1957

[26] *The Interaction Lagrangian for Fundamental Particles*, Nucl. Phys. **2**, 173–176.

[27] (with P. T. Matthews) *The Coupling Constants of Strong Interactions*, Nucl. Phys. **2**, 582–585.

[28] *On Parity Conservation and Neutrino Mass*, Nuovo Cimento **5**, 299–301.

[29] (with B. D'Espagnat and J. Prentki) *On Symmetries in Elementary Particle Interactions*, Nucl. Phys. **3**, 446–455.

[30] *On Fermi Interactions*, Imperial College preprint.

1958

[32] (with B. D'Espagnat and J. Prentki) *On Symmetries of Weak Interactions*, Nucl. Phys. **5**, 447–454.

[31] *On the Inter-Relation of Scalar and Pseudoscalar Theories*, Nucl. Phys. **5**, 687–689.

[33] (with P. T. Matthews) *An Analysis of the Strong Interaction Data*, Nuovo Cimento **7**, 789–793.

[34] *Elementary Particles*, Pakistan J. Sci. **10**, No. 2.

[35] (with P. T. Matthews) *K-Meson Dispersion Relations: Part I - Theory*, Phys. Rev. **110**, 565–568.

[36] (with P. T. Matthews) *K-Meson Dispersion Relations: Part II - Applications*, Phys. Rev. **110**, 569–575.

[37] (with P. T. Matthews) *Relativistic Field Theory of Unstable Particles*, Phys. Rev. **112**, 283–287.

[37a] *Analytic Properties of Expectation Values of Products of Field Operations*, **AT30**, 875.

1959

[38] (with J. Tiomno) *On the Masses of Elementary Particles*, Nucl. Phys. **9**, 585–587.

[39] (with J. C. Ward) *Weak and Electromagnetic Interactions*, Nuovo Cimento **11**, 568–577.

[40] (with P. T. Matthews) *Evidence for Resonance States in the K^--N System*, Phys. Rev. Lett. **2**, 226–227.

[41] (with P. T. Matthews) *The Inelastic Scattering of Elementary Particles*, Nuovo Cimento **13**, 381–393.

[42] (with P. T. Matthews) *Relativistic Theory of Unstable Particles II*, Phys. Rev. **115**, 1079–1084.

1960

[43] *Invariance Properties in Elementary Particle Physics*, in Lectures in Theoretical Physics, Vol. 2 (Interscience, New York), 1–30.

[44] (with J. C. Polkinghorne) *On the Possible Existence of Doubly Strange Heavy Mesons*, Nuovo Cimento **15**, 166–167.

[45] (with G. Feldman and P. T. Matthews) *The Inelastic Scattering of Elementary Particles II*, Nuovo Cimento **16**, 549–555.

[46] *An Equivalence Theorem for Partially Gauge-Invariant Vector Meson Interactions*, Nucl. Phys. **18**, 681–690.

[47] (with A. Komar) *Renormalization Problem for Vector Meson Theories*, Nucl. Phys. **21**, 624–630.

[48] (with J. C. Ward) $\Delta I = 1/2$ *Rule*, Phys. Rev. Lett. **5**, 390.

1961

[49] (with J. C. Ward) *On a Gauge Theory of Elementary Interactions*, Nuovo Cimento **19**, 165–170.

[50] (with G. Feldman and P. T. Matthews) *Nonleptonic Decay Modes of the Hyperons*, Phys. Rev. **121**, 302–303.

[51] (with P. T. Matthews) *Resonance Effects in Intermediate Boson Theory*, Nuovo Cimento **19**, 837–839.

[52] (with J. C. Ward) *Vector Field Associated with the Unitary Theory of the Sakata Model*, Nuovo Cimento **20**, 419–421.

[53] (with S. Kamefuchi and L. O'Raifeartaigh) *Change of Variables and Equivalence Theorems in Quantum Field Theories*, Nucl. Phys. **28**, 529–549.

[54] (with J. C. Ward) *On the Symplectic Symmetry*, Nuovo Cimento **20**, 1228–1230.

[55] (with P. T. Matthews) *Diffraction Scattering of Elementary Particles*, Nuovo Cimento **21**, 126–134.

[56] *Some Speculations on the New Resonances*, Rev. Mod. Phys. **33**, 426–430.

[57] (with P. T. Matthews) *Dispersion Relations and Diffraction Scattering*, Nuovo Cimento **21**, 823–828.

[58] (with J. Tiomno) *Isotopic Spin Relations in Hyperon Production*, Nuovo Cimento **22**, 1287–1289.

1962

[59] *Some Consequences of the Unitary Symmetry*, Nuovo Cimento **23**, 455–457.

[61] *Elementary Particle Resonances*, Proc. Phys. Soc. **80**, 13–27.

[60] (with P. T. Matthews) *Unitary Symmetry*, Proc. Phys. Soc. **80**, 28–38.

[62] *Renormalizability of Gauge Theories*, Phys. Rev. **127**, 331–334.

[63] *Lagrangian Theory of Composite Particles*, Nuovo Cimento **25**, 224–227.

[64] (with J. Goldstone and S. Weinberg) *Broken Symmetries*, Phys. Rev. **127**, 965–970.

1963

[65] *Renormalizable Electrodynamics of Vector Mesons*, Phys. Rev. **130**, 1287–1290.

[66] *The Formalism of Lie Groups*, in Theoretical Physics (IAEA, Vienna).

1964

[67] *Universal Weak Interaction Lagrangian*, Phys. Lett. **8**, 216–217.

[68] (with P. T. Matthews) *A Remarkable Representation of the Weak and Electromagnetic Lagrangian*, Phys. Lett. **8**, 357–358.

[69] (with P. T. Matthews) *K_1-K_2 Transition in a Coulomb Field*, Nuovo Cimento **32**, 778.

[73] (with R. Delbourgo) *Renormalizable Electrodynamics of Scalar and Vector Mesons: Part II*, Phys. Rev. **135**, B1398–B1427.

[70] *Gauge Theory of Elementary Interactions* Phys. Rev. **136**, B763–B768.

[72] (with J. C. Ward) *Electromagnetic and Weak Interactions*, Phys. Lett. **13**, 168–171.

[74] *Symmetries of Strong Interactions*, Proc. XIIth Int. Conf. on High Energy Physics (Atomizdat, Moscow), 799–812.

[75] *On the Algebra of SU(6)*, Phys. Lett. **13**, 354–355.

[76] *Summary of Conference Results*, Nucleon Structure (Stanford University Press).

[77] *Developments in Renormalization Theory*, Analysis in Function Space (MIT, Cambridge, MA), Chap. 7.

1965

[78] (with R. Delbourgo and J. Strathdee) *Relativistic Structure of SU(6)*, Phys. Rev. **138**, B420–B423.

[79] (with R. Delbourgo and J. Strathdee) *$\tilde{U}(12)$ and Broken SU(6) Symmetry*, Nuovo Cimento **36**, 689–692.

[80] (with R. Delbourgo and J. Strathdee) *The Covariant Theory of Strong Interaction Symmetries*, Proc. R. Soc. (London) **A284**, 146–158.

[81] (with R. Delbourgo, M. A. Rashid and J. Strathdee) *The Covariant Theory of Strong Interaction Symmetries II*, Proc. R. Soc. (London) **A285**, 312–318.

[82] (with J. M. Charap, P. T. Matthews and J. Strathdee) *P and C Properties of the $\tilde{U}(12)$ Multiplets*, Phys. Lett. **15**, 184–185.

[83] (with R. Delbourgo, M. A. Rashid and J. Strathdee) *The $\tilde{U}(12)$ Symmetry*, in High Energy Physics and Elementary Particles (IAEA, Vienna), 455–532.

[84] (with R. Delbourgo and J. Strathdee) *Relativistic Extension of Non-Compact Symmetry Groups*, Proc. R. Soc. (London) **A289**, 177–191.

[85] *The Role of Symmetry Physics: Some Conclusions from the Oxford Conference, September 1965*, Proc. Int. Conference on Elementary Particles, Oxford.

1966

[87] *Theory of Groups and the Symmetry Physicist*, J. London Math. Soc. **41**, 49–62.

[86] (with J. Strathdee) *On the Coupling of Non-Compact Group Representations*, Proc. R. Soc. (London) **A292**, 314–321.

[88] (with J. Strathdee) *Relativistic U(6,6) Theory*, Phys. Rev. **148**, 1352–1358.

[89] (with R. Delbourgo and J. Strathdee) *On the Dynamical Origin of SU(6)*, Phys. Lett. **21**, 455–457.

[91] (with R. Delbourgo and J. Strathdee) *Multiple Field Relations*, Phys. Lett. **22**, 680–683.

[91a] *Magnetic Monopole and Two-Photon Theories of C-Violation*, Phys. Lett. **22**, 683–684.

1967

[90] (with R. Delbourgo and J. Strathdee) *Bound-State Equation for Quark-Antiquark Systems*, Nuovo Cimento **50**, 193–210.

[92] (with R. Delbourgo and J. Strathdee) *Stress Tensor and the 2^+ Mesons*, Nuovo Cimento **49**, 593–605.

[93] (with R. Delbourgo and J. Strathdee) *A Convergent Field Theory*, Phys. Lett. **24B**, 622.

[95] (with J. F. Boyce, R. Delbourgo and J. Strathdee) *Partial Wave Analysis Part I*, ICTP, Trieste preprint IC/67/9.

[94] (with R. Delbourgo and J. Strathdee) *Harmonic Analysis in Terms of the Homogeneous Lorentz Group*, Phys. Lett. **25B**, 230–232.

[96] (with R. Delbourgo and J. Strathdee) *Partial Wave Analysis in Terms of the Homogeneous Lorentz Group*, Phys. Rev. **164**, 1981–1990.

[97] (with J. Strathdee) *Reggeization of Internal Symmetries*, Phys. Rev. Lett. **19**, 339–343.

[98] (with R. Delbourgo, M. A. Rashid and J. Strathdee) *Green's Functions for Regge Trajectories*, Phys. Lett. **25B**, 475–476.

[98a] (with R. Delbourgo, M. A. Rashid and J. Strathdee) *Reggeization of Internal Symmetries II*, ICTP Internal Report 21/1967.

1968

[102] (with R. Delbourgo, M. A. Rashid and J. Strathdee) *Reggeization of U(6) × U(6)*, Phys. Rev. **170**, 1477–1489.

[99] (with R. Delbourgo and J. Strathdee) *Reggeization of Quark Number*, Phys. Rev. **172**, 1727–1736.

[100] (with J. Strathdee) *On the Zero Energy Behaviour of Regge Poles and Residues*, Nobel Symp. 8, Elementary Particle Theory-Relativistic Groups and Analyticity (Wiley), 59–103.

[101] *Weak and Electromagnetic Interactions*, Nobel Symp. 8, Elementary Particle Theory-Relativistic Groups and Analyticity (Wiley), 367–377.

[103] (with R. Delbourgo and J. Strathdee) *A Yang-Mills Gauge Theory of Supermultiplets*, ICTP, Trieste preprint IC/68/5.

1969

[104] (with G. Mark) *Finite-Component Field Representations of the Conformal Group*, Ann. Phys. **53**, 174–202.

[105] *Fundamental Theory of Matter: A Survey of Results and Methods*, in Contemporary Physics, Vol. II (IAEA, Vienna).

[106] (with J. Strathdee) *Nonlinear Realizations: I. The Role of Goldstone Bosons*, Phys. Rev. **184**, 1750–1759.

[107] (with J. Strathdee) *Nonlinear Realizations: II. Conformal Symmetry*, Phys. Rev. **184**, 1760–1768.

[108] (with R. Delbourgo) *Reggeization in Supermultiplet Theories*, Phys. Rev. **186**, 1516–1528.

[109] (with R. Delbourgo) *Analysis of Charge Exchange Meson-Baryon Reactions from Reggeized Supermultiplet Theory*, Phys. Lett. **28B**, 497–500.

[110] (with R. Delbourgo and J. Strathdee) *Infinities of Nonlinear and Lagrangian Theories*, Phys. Rev. **187**, 1999–2007.

[111] (with R. Delbourgo and J. Strathdee) *Suppression of Infinities in Einstein's Gravitational Theory*, Nuovo Cimento **2**, 354–359.

[112] (with R. Delbourgo, A. P. Hunt and Q. Shafi) *Self-Momentum in Chiral and Gravitational Theories*, ICTP, Trieste preprint IC/69/15.

1970

[113] (with J. Strathdee) *Momentum-Space Behavior of Integrals in Nonpolynomial Lagrangian Theories*, Phys. Rev. **D1**, 3296–3312.

[114] (with C. Isham and J. Strathdee) *Spontaneous Breakdown of Conformal Symmetry*, Phys. Lett. **31B**, 300–302.

[116] *Nonpolynomial Lagrangian Theories*, Acta Physica Austriaca Supp. VII, 1–31.

[118] (with C. Isham and J. Strathdee) *Broken Chiral and Conformal Symmetry in an Effective Lagrangian Formalism*, Phys. Rev. **D2**, 685–690.

[117] (with J. Strathdee) *Equivalent Formulations of Massive Vector Field Theories*, Phys. Rev. **D2**, 2869–2876.

[119] (with J. Strathdee) *Quantum Gravity and Infinities in Quantum Electrodynamics*, Nuovo Cimento Lett. **4**, 101–108.

[120] *On Renormalization Constants and Inter-Relation of Fundamental Forces.* Lecture at Kiev Conference.

[121a] (with C. J. Isham and J. Strathdee) *Non-Polynomial Interactions and the K-Matrix*, ICTP, Trieste preprint IC/70/12.

1971

[125] (with C. Isham and J. Strathdee) *Nonlinear Realizations of Space-Time Symmetries: Scalar and Tensor Gravity*, Ann. Phys. **62**, 98–119.

[130] (with R. Delbourgo and K. Koller) *Renormalization of Rational Lagrangians*, Ann. Phys. (N. Y.) **66**, 1–19.

[123] (with C. Isham and J. Strathdee) *f-Dominance of Gravity*, Phys. Rev. **D3**, 867–873.

[122] (with C. Isham and J. Strathdee) *Infinity Suppression in Gravity-Modified Quantum Electrodynamics*, Phys. Rev. **D3**, 1805–1817.

[126] (with J. Strathdee) *Gravitational Ward Identities* (dedicated to Prof. Igor Tamm), ICTP, Trieste internal report IC/71/1.

[127] *Computation of Renormalization Constants*, in Renormalization and Gravity (Gordon and Breach).

[128] (with C. Isham and J. Strathdee) *Gauge-Invariant Infinity Suppression in Gravity Modified Quantum Electrodynamics*, Phys. Lett. **35B**, 585–588.

[129] *Finite Field Theories*, in Particles and Fields — 71 (AIP, New York), 259–274.

[134] *Symmetry Concepts and the Fundamental Theory of Matter*, J. Sci. **1**, 1–15.

1972

[124] (with R. Delbourgo) *The Gravitational Correction to PCAC*, Phys. Lett. **40B**, 381–387.

[131] (with C. Isham and J. Strathdee) *Infinity Suppression in Gravity-Modified Electrodynamics: II*, Phys. Rev. **D5**, 2548–2565.

[132] (with J. Strathdee) *A Renormalizable Gauge Model of Lepton Interactions*, Nuovo Cimento **11A**, 397–435.

[133] (with J. Strathdee) *The Path-Integral Quantization of Gravity*, Aspects of Quantum Theory (Cambridge University Press), 249–268.

[135] *Nonpolynomial Lagrangians, Infinities and Gravity Theory*, in Developments in High Energy Physics (Academic Press, New York), 386–422.

[136] (with C. Isham and J. Strathdee) *CP-Violating Gauge Interactions*, Nuovo Cimento Lett. **4**, 678.

[137] (with R. Delbourgo) *PCAC Anomalies and Gravitation*, ICTP, Trieste preprint IC/72/86.

[138] (with C. Isham and J. Strathdee) *SL(6,C) Gauge Invariance of Einstein-Like Lagrangians*, Nuovo Cimento Lett. **5**, 969.

1973

[121] *Breaking of Conformal Symmetry and Renormalization of Non-Polynomial Theories*, in The Past Decade in Particle Theory (Gordon and Breach, London), 1–32.

[152] (with J. C. Pati) *Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions*, Phys. Rev. **D8**, 1240–1251.

[139] (with C. Isham and J. Strathdee) *The 2^+ Nonet as Gauge Particles for SL(6,C) Symmetry*, Phys. Rev. **D8**, 2600–2609.

[140] (with R. Delbourgo) *The Mass Matrix for a Leptonic U(3) Model*, ICTP, Trieste preprint IC/73/1.

[141] *Progress in Renormalization Theory Since 1949*, Lecture at Development of the Physicist's Conception of Nature (Dirac's 70th Birthday) ICTP, Trieste preprint IC/73/2.

[142] (with C. Isham and J. Strathdee) *Vector and Tensor Gauge Particles in SL(6,C) Theory*, ICTP, Trieste preprint IC/73/22.

[143] *On SL(6,C) Gauge Invariance*, Lecture at First Decade of Coral Gables Conference on Fundamental Interactions, January, ICTP, Trieste preprint IC/73/28.

[145] (with C. Isham and J. Strathdee) *The Influence of f Gravity on Gravitational Collapse*, Nature Phys. Sci. **244**, 82–84.

[147] (with C. Isham and J. Strathdee) *Wilson Operator Products in Nonlinearly Realized Theories*, ICTP, Trieste preprint IC/73/54.

[151] (with J. C. Pati) *A U(3) Nonet Scheme for Leptons*, Phys. Lett. **43B**, 311–313.

[148] (with C. Isham and J. Strathdee) *Is Quantum Gravity Ambiguity-Free?*, Phys. Lett. **46B**, 407–411.

[149] (with J. C. Pati) *Inconsistency of a Class of Gauge Theories Based on Han-Nambu-like Quarks*, ICTP, Trieste internal report IC/73/81.

[150] (with J. C. Pati) *Is Baryon Number Conserved?*, Phys. Rev. Lett. **31**, 661–664.

1974

[141a] (with R. Delbourgo and J. Strathdee) *Scalar Multiplets and Asymptotic Freedom*, Nuovo Cimento **23A**, 237–256.

[144] (with J. Strathdee) *Comment on the Computation of Effective Potentials*, Phys. Rev. **D9**, 1129–1131.

[146] (with C. Isham and J. Strathdee) *An f-g Version of the SL(6,C) Gauge Theory*, Phys. Rev. **D9**, 1702–1705.

[153] (with J. Strathdee) *Coupling Constants are not Arbitrary in Asymptotically Free Theories*, Nuovo Cimento Lett. **9**, 464–466.

[153a] (with J. Strathdee) *Sum Rules in Asymptotically Free Theories*, Nuovo Cimento Lett. **10**, 113–114.

[154] (with J. C. Pati) *Are There Anomalous Lepton-Hadron Interactions?*, Phys. Rev. Lett. **32**, 1083–1086.

[155] (with J. C. Pati) *Lepton Number as the Fourth "Color"*, Phys. Rev. **D10**, 275–289.

[156] (with J. Strathdee) *Super-Gauge Transformations*, Nucl. Phys. **B76**, 477–482.

[157] (with J. Strathdee) *Unitary Representations of Super-Gauge Symmetries*, Nucl. Phys. **B80**, 499–505.

[158] (with J. Strathdee) *On Goldstone Fermions*, Phys. Lett. **49B**, 465–467.

[159] (with J. Strathdee) *Supersymmetry and Non-Abelian Gauges*, Phys. Lett. **51B**, 353–355.

[161] (with R. Delbourgo and J. Strathdee) *Supersymmetric V-A Gauges and Fermion Number*, Phys. Lett. **51B**, 475–478.

[162] *Impact of Quantum Gravity Theory on Particle Physics*, Lecture at Oxford Conference on General Relativity and Particle Physics, February, ICTP, Trieste preprint IC/74/55.

[166] *Are Leptons Really Different from Hadrons?*, Lecture at Erice Summer School, July, ICTP, Trieste preprint IC/74/87.

[169] (with J. Strathdee) *Transition to CP Conservation and Zero Cabibbo Angle in Strong Magnetic Fields*, Nature **252**, 569–571.

1975

[160] (with J. Strathdee) *Superfields and Fermi-Bose Symmetry*, Phys. Rev. **D11**, 1521–1535.

[163] (with J. Strathdee) *SU(6) and Supersymmetry*, Nucl. Phys. **B84**, 127–131.

[164] (with J. C. Pati) *Anomalous Lepton-Hadron Interactions and Gauge Models*, Phys. Rev. **D11**, 1137–1154.

[165] (with J. Strathdee) *Feynman Rules for Superfields*, Nucl. Phys. **B86**, 142-152.

[167] (with J. C. Pati and J. Strathdee) *On Fermion-Number and Its Conservation*, Nuovo Cimento **26A**, 72–83.

[168] *Supersymmetry and Fermion-Number Conservation*, Nucl. Phys. **B87**, 85–92.

[170] (with J. Strathdee) *Transition Electromagnetic Fields in Particle Physics*, Nucl. Phys. **B90**, 203–220.

[171] (with J. C. Pati) *Are the New Particles Color Gluons?*, Phys. Rev. Lett. **34**, 613–616.

[172] (with J. C. Pati) *Color and Charm for the y-Particles*, University of Maryland Technical Report 75-056.

[173] (with J. Strathdee) *A Theorem Concerning Goldstone Fermions*, Lett. Math. Phys. **1**, 3–6.

[174] (with J. Strathdee) *A Supersymmetric Gluon Model*, Nucl. Phys. **B93**, 23–28.

[175] (with J. Strathdee) *Supersymmetry, Parity and Fermion Number Conservation*, Nucl. Phys. **B97**, 293–316.

[176] (with T. W. B. Kibble and J. Strathdee) *Intensity-Dependent Mass Shift and Symmetry Breaking*, Nucl. Phys. **B96**, 255–263.

[178] (with J. C. Pati) *Mirror Fermions, J/ψ Particles, Kolar-Mine Events and Neutrino Anomaly*, Phys. Lett. **58B**, 333–337.

[180] (with J. C. Pati) *Quarks, Leptons and Pre-Quarks*, Palermo Conference, June, ICTP, Trieste preprint IC/75/106.

[181] (with R. Delbourgo) *The Stueckelberg Formalism for Spin Two*, Nuovo Cimento Lett. **12**, 297–299.

[182] (with J. C. Pati and J. Strathdee) *Are Quarks Composite?*, Phys. Lett. **59B**, 265–268.

[183] *Unification of Fundamental Forces and Elementary Particle Physics*, Lecture at 1st Scientific Conference, Iraqi Atomic Energy Commission, Baghdad, April.

1976

[179] (with J. C. Pati) *Why Color Fails to Show in Electroproduction and Neutrino Scattering Experiments*, Phys. Rev. Lett. **36**, 11–14.

[177] (with R. N. Mohapatra and J. C. Pati) *'Color' as a Classification Symmetry*, Phys. Rev. **D13**, 1733–1743.

[185] (with J. C. Pati and S. Sakakibara) *Are SLAC (μe) Events Decay Products of Unconfined Integer Charge Quarks?*, Phys. Rev. Lett. **36**, 1229–1233.

[186] (with D. M. Capper and J. Strathdee) *A Massless Non-Goldstone Neutrino in a Supersymmetric Model*, Nucl. Phys. **B107**, 137–145.

[187] (with J. Strathdee) *Black Holes as Solitons*, Phys. Lett. **61B**, 375–376.

[188] (with J. Strathdee) *Mass Problem for Tensor Mesons*, Phys. Rev. **D14**, 2830–2832.

[189] *The Unconfined Unstable Quark (Predictions from a Unified Gauge Theory of Strong, Weak and Electromagnetic Interaction)*, Lecture at CERN, February, ICTP, Trieste preprint IC/76/21.

[190] (with J. Strathdee) *Symmetry, Order and Symmetry Restoration*, Lecture at Brookhaven National Laboratory, June, ICTP, Trieste preprint IC/76/62.

[191] (with J. C. Pati) *New Sources of Multileptons in 1N and e^-/e^+ Ratio in NN Collisions*, ICTP, Trieste preprint IC/76/63.

[192] (with J. C. Pati) *Lepton-Hadron Unification*, ICTP, Trieste preprint IC/76/76.

1977

[195c] (with J. C. Pati and S. Rajpoot) *Natural Left-Right Symmetry, Atomic Parity Experiments and Asymmetries in High Energy e^+e^- Collisions in PETRA and PEP regions*, Phys. Lett. **71B**, 387–391.

[184] *Probing into the Heart of Matter*, in Eucyclopaedia of Ignorance (Pergamon).

[198] (with J. C. Pati) *Nucleon Dissociation and Quark and Gluon Decays as New Sources of Multi-Lepton Events*, Phys. Rev. **D15**, 1406–1411.

[193] (with J. Strathdee) *Hadronic Temperature and Black Solitons*, Phys. Lett. **66B**, 143–146.

[194] (with V. Elias, J. C. Pati and J. Strathdee) *Charged Spin-1 Gluons, Parton Model and the Archimedes Effect*, Pramana **8**, 303–308.

[195] (with J. Strathdee) *Class of Solutions for the Strong Gravity Equations*, Phys. Rev. **D16**, 2668–2672.

[196] *Gauge Unification of Basic Forces Particularly of Gravitation with Strong Interactions*, Lecture at R.E. Marshak's 60th Birthday, January, Ann. New York Acad. Sci. **294**, 12–36.

[197] (with J. Strathdee) *A Steeply-Rising Potential in Tensor Gauge Theory*, Phys. Lett. **67B**, 429–431.

[200] (with J. C. Pati) *Design of Future Experiments*, ICTP, Trieste preprint IC/77/65.

[202] (with H. S. Mani, J. C. Pati, and S. Rajpoot) *Dilepton Production in PP and PP Collisions as a Probe to the Nature of the Neutral Current Interaction*, Phys. Lett. **72B**, 75–79.

1978

[204] (with V. Elias and J. C. Pati) *Unification of the Basic Particle Forces at a Mass-Scale of Order 1000 m_W*, University of Maryland technical report 78-041.

[195a] (with J. C. Pati and S. Rajpoot) *Neutral-Current Phenomena within the Left-Right-Symmetric Unified Theory of Quarks and Leptons*, Phys. Rev. **D17**, 131–150.

[187a] (with J. Strathdee) *Supersymmetry and Superfields*, Fortsch. Phys. **26**, 57–142.

[199] (with J. Strathdee) *Classical Confining Solutions of a Tensor Gauge Theory Incorporating Color*, Phys. Rev. **D18**, 1323–1326.

[201] (with H. S. Mani and J. C. Pati) *"Naturalness" of Atomic Parity Conservation within Left-Right Symmetric Unified Theories*, Phys. Rev. **D17**, 2510–2514.

[203] (with A. H. Chamseddine and J. Strathdee) *Strong Gravity and Supersymmetry*, Nucl. Phys. **B136**, 248–258.

[205] (with J. Strathdee) *Confinement through Tensor Gauge Fields*, Phys. Rev. **D18**, 4596–4609.

[206] (with J. Strathdee) *Remarks on High-Energy Stability and Renormalizability of Gravity Theory*, Phys. Rev. **D18**, 4480–4485.

[207] (with J. Strathdee) *Higgs Couplings and Asymptotic Freedom*, Phys. Rev. **D18**, 4713–4720.

[208] (with J. C. Pati) *Liberated Quarks and Gluons in $e^- e^+$ Jets and Beam-Dump Experiments*, Nucl. Phys. **B144**, 445–461.

[209] (with V. Elias and J. C. Pati) *Light Neutral Z-Boson in the PETRA and PEP Energy Range within the Unifying Symmetry $[SU(4)]^4$*, Phys. Lett. **73B**, 451–456.

[212] (with J. C. Pati, S. Rajpoot and J. Strathdee) *Evidence for Leptonic Color?* Nature **273**, 736–738.

[211a] *Unification, Superunification and New Theoretical Ideas*, Proc. 19th Int. Conf. High Energy Physics (Physical Society of Japan, Tokyo) 933–957.

1979

[210] (with J. C. Pati) *Signatures for Axial Chromodynamics*, Nucl. Phys. **B150**, 76–86.

[211] *The Electroweak Force, Grand Unification and Superunification*, Phys. Scripta **20**, 216–226.

[213] (with N. S. Baaklini) *Non-Metrical Supergravity for a Spin-Two Nonet*, Nucl. Phys. **B153**, 237–243.

[214] (with N. S. Craigie) *On the Effect of Scalar Partons at Short Distances in Unified Theories with Spontaneously Broken Color Symmetry*, Phys. Lett. **85B**, 57–62.

[215] *A Gauge Appreciation of Developments in Particle Physics — 1979*, ICTP, Trieste preprint IC/79/142, Talk at EPS Conference, Geneva.

1980

[195b] (with J. C. Pati and S. Rajpoot) *Pre-Quarks and Fractional Charges*, J. Phys. A: Math.-Gen. **13**, 2509–2512.

[216] *Gauge Unification of Fundamental Forces*, Rev. Mod. Phys. **52**, 525–538.

[217] (with V. Elias) *Induced Higgs Couplings and Spontaneous Symmetry Breaking*, Phys. Rev. **D22**, 1469–1472.

[218] *The Nature of the "Ultimate" Explanation in Physics*, talk given at Herbert Spencer Lectures, Oxford, 12 November, ICTP, Trieste preprint IC/80/39.

[219] (with J. Strathdee) *Fermion Masses and the Gauge Hierarchy Problem*, ICTP, Trieste preprint IC/80/67.

[220] (with J. C. Pati) *Quark-Lepton Unification and Proton Decay*, ICTP, Trieste preprint, IC/80/72.

[221] (with J. Strathdee) *On Witten's Charge Formula*, Lett. Math. Phys. **4**, 505–508.

[222] (with J. Strathdee) *Dynamical Mass Generation in* $(U_L(1) \times U_R(1))^2$, ICTP, Trieste internal report IC/80/95.

1981

[223] *A Preon Model with Hidden Electric and Magnetic Type Charges*, Nucl. Phys. **B185**, 416–428.

[224] (with J. Pati and J. Strathdee) *Probings through Proton Decay and n-n Oscillations*, Nucl. Phys. **B185**, 445–472.

[226] (with J. Strathdee) *On the Montenon-Olive Conjecture*, ICTP, Trieste internal report IC/81/81.

[227] *Proton-Decay as a Window on Highest Energy Physics*, Proc. Int. Conference on High Energy Physics, Lisbon, July, (EPS), 380–389.

[228] (with J. Strathdee) *On Kaluza-Klein Theory*, Ann. Phys. **141**, 316–352.

[229] (with J. C. Pati and J. Strathdee) *Preons and Supersymmetry*, To honour Francis Low's Sixtieth Birthday, ICTP, Trieste internal report IC/81/222.

[230] *Developments in Fundamental Physics*, Lecture given at the EPS Conference, Istanbul, ICTP, Trieste internal report, IC/81/223.

1982

[225] *Gauge Interactions, Elementarity and Superunification*, Phil. Trans. R. Soc. **304**, 135–153.

[231] (with J. Strathdee and C. Panagiotakopoulos) *Supersymmetric Field Theory of Monopoles*, Phys. Lett. **115B**, 29–32.

[232] (with J. C. Pati and J. Strathdee) *Can SU(16) be Ruled Out by Low Energy Experiments?*, Phys. Lett. **108B**, 121–126.

[233] (with E. Sezgin) *Maximal Extended Supergravity Theory in Seven Dimensions*, Phys. Lett. **118B**, 359–364.

[236] *Physics with 100–1000 TeV Accelerators*, ICTP, Trieste preprint IC/82/215.

1983

[234] (with J. C. Pati) *Approach Towards Minimality — Supersymmetry at the Preonic or Pre-Preonic Level*, Nucl. Phys. **B214**, 491–512.

[235] (with S. Randjbar-Daemi and J. Strathdee) *Spontaneous Compactification in Six-Dimensional Einstein-Maxwell Theory*, Nucl. Phys. **B214**, 491–512.

[237] (with M. A. Namazie and J. Strathdee) *Finiteness of Broken N=4 Super Yang-Mills Theory*, Phys. Rev. **D28**, 1481–1490.

[238] (with S. Randjbar-Daemi and J. Strathdee) *Instability of Higher-Dimensional Yang-Mills Systems*, Phys. Lett. **124B**, 345–348.

[239] (with S. Randjbar-Daemi and J. Strathdee) *Compactification of Supergravity Plus Yang-Mills in 10 Dimensions*, Phys. Lett. **124B**, 349–352.

[240] (with E. Sezgin) *SO(4) Gauging of N=2 Supergravity in Seven Dimensions*, Phys. Lett. **126B**, 295–300.

[241] (with S. Randjbar-Daemi and J. Strathdee) *Instanton Induced Compactification and Fermion Chirality*, Phys. Lett. **132B**, 56–60.

[243] (with J. C. Pati and Utpal Sarkar) *Decay Modes in* $SU(2)_L \times SU(2)_R \times SU(4)_{\rm col}$ *or SO(10)*, Phys. Lett. **133B**, 330–336.

1984

[242] (with J. C. Pati) *Supersymmetric Flavon-Chromon Models*, Nucl. Phys. **B234**, 223–227.

[244] (with S. Randjbar-Daemi and J. Strathdee) *On Kaluza-Klein Cosmology*, Phys. Lett. **135B**, 388–392.

[245] (with S. Randjbar-Daemi and J. Strathdee) *Stability of Instanton Induced Compactification in Eight Dimensions*, Nucl. Phys. **B242**, 447–472.

[247] (with S. Randjbar-Daemi and J. Strathdee) *Towards a Self-Consistent Computation on Vacuum Energy in 11-dimensional Supergravity*, Nuovo Cimento **84B**, 167–180.

[248] (with E. Sezgin) *Chiral Compactification on Minkowski* $\times S^2$ *of* $N = 2$ *Einstein-Maxwell Supergravity in Six Dimensions*, Phys. Lett. **147B**, 47–51.

[248a] (with S. Randjbar-Daemi and J. Strathdee) *Chiral Families and Stable Compactifications*, ICTP, Trieste internal report IC/84/62.

[249] *Kaluza-Klein Proposal and Electro-Nuclear Gravity*, ICTP, Trieste preprint IC/84/170, Report to the Leipzig Conference on High Energy Physics, July.

1985

[250] (with E. Sezgin) $d = 8$ *Supergravity*, Nucl. Phys. **B258**, 284–304.

[251] (with S. Randjbar-Daemi, E. Sezgin and J. Strathdee) *An Anomaly-Free Model in Six Dimensions*, Phys. Lett. **151B**, 351–356.

[252] (with E. Sezgin) *$d = 8$ Supergravity: Matter Couplings, Gauging and Minkowski Compactification*, Phys. Lett. **154B**, 37–42.

[253] (with E. Sezgin) *Anomaly Freedom in Chiral Supergravities*, in honour of Stig Lundqvist's 60th birthday, Phys. Scripta **32**, 283–285.

[257] (with S. Randjbar-Daemi and J. Strathdee) *Remarks on the Gauge Fixing Problems of the Spinning String in Neveu-Schwarz-Ramond Formalism*, ICTP, Trieste internal report IC/85/311.

1986

[254] (with E. Bergshoeff, S. Randjbar-Daemi, H. Sarmadi and E. Sezgin) *Locally Supersymmetric σ-Model with Wess-Zumino Term in Two Dimensions and Critical Dimensions for Strings*, Nucl. Phys. **B269**, 77–96.

[255] *Astroparticle Physics*, in Proc. of the 4th Marcel Grossmann Meeting on General Relativity (Elsevier).

[256] (with J. C. Pati and B. V. Sreekantan) *The Moon as the Searching Ground for Proton Decay*, Int. J. Mod. Phys. **A1**, 147–153.

[258] *Overview of Particle Physics*, ICTP, Trieste preprint IC/86/35.

[259] (with E. Bergshoeff and E. Sezgin) *A Supersymmetric R^2-Action in Six Dimensions and Torsion*, Phys. Lett. **B173**, 73–76.

[261] (with E. Bergshoeff, T. W. Kephart and E. Sezgin) *Global Anomalies in Six Dimensions*, Mod. Phys. Lett. **A1**, 267–276.

1987

[260] (with E. Bergshoeff and E. Sezgin) *Supersymmetric R^2 Actions, Conformal Invariance and Lorentz Chern-Simons Term in 6 and 10 dimensions*, Nucl. Phys. **B279**, 659–683.

[246] (with S. Randjbar-Daemi and J. Strathdee) *On $SU(3) \times SU(2) \times U(1)$ Invariant Compactifying Solutions to 11-Dimensional Supergravity*, in Quantum Field Theory and Quantum Statistics, Vol. 2, 217–232.

[262] (with S. Randjbar-Daemi and J. Strathdee) *Conformal Anomaly in a Covariantly Quantized Heterotic Green-Schwarz Superstring*, Mod. Phys. Lett. **A2**, 145–150.

[263] (with S. Randjbar-Daemi and J. Strathdee) *σ-Models and String Theories*, Int. J. Mod. Phys. **A2**, 667–693.

1988

[264] *Particle Physics Today (1987)*, Ann. Inst. Henri Poincare **49**, 369–385.

[265] (with E. Bergshoeff, E. Sezgin and Y. Tanii) *Singletons, Higher Spin Massless States and the Supermembrane*, Phys. Lett. **B205**, 237–244.

[266] (with E. Bergshoeff, E. Sezgin and Y. Tanii) *$N = 8$ Supersingleton Quantum Field Theory*, Nucl. Phys. **B305**, 497–515.

1989

[267] *Astroparticle Physics (1988)*, Int. J. Mod. Phys. **A4**, 583–605.

[268] (with S. Randjbar-Daemi and J. Strathdee) *On Heterotic Supermanifolds*, Nucl. Phys. **B320**, 221–252.

1990

[269] (with S. Randjbar-Daemi and J. Strathdee) *Chern-Simons Superconductivity at Finite Temperature*, Nucl. Phys. **B340**, 403–447.

[270] (with S. Randjbar-Daemi and J. Strathdee) *Anyons and Chern-Simons Theory on Compact Spaces of Finite Genus*, Phys. Lett. **B240**, 121–126.

1991

[271] *The Role of Chirality in the Origin of Life*, J. Mol. Evol. **33**, 105–113.

[272] (with S. Randjbar-Daemi and J. Strathdee) *On Charged Fermions in Two Dimensions*, Int. J. Mod. Phys. **B5**, 845–866.

[273] (with S. Randjbar-Daemi and J. Strathdee) *Fermion Induced Singularities in a Non-Relativistic Effective Lagrangian for Electromagnetism*, Int. J. Mod. Phys. **B5**, 1419–1436.

1992

[274] *Chirality, Phase Transitions and Their Induction in Amino Acids*, Phys. Lett. **B288**, 153–160.

1993

[275] (with S. Randjbar-Daemi and J. Strathdee) *Generalized Spin Systems and σ-Models*, Phys. Rev. **B48**, 3190–3205.

[276] (with C. Sivaram) *Strong Gravity Approach to QCD and Confinement*, Mod. Phys. Lett. **A8**, 321–326.

2265 hc

ISBN 981-02-1662-9